



```

; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 960 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...957
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-121-032-60

Query Match      100.0%; Score 960; DB 14; Length 960;
Best Local Similarity 100.0%; Pred. No. 3.4e-272;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAAGTTGGTATCTATCTTAAACAATCTTTTATAGCAGCAATA 60
DB 1 ATGAGCAAGAAAAAGTTGGTATCTATCTTAAACAATCTTTTATAGCAGCAATA 60
QY 61 TATTTTGTAGAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
DB 61 TATTTTGTAGAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
QY 121 CCACCCCAACACACTTTCCTACTACCAAGTTCCTCAAGATTAGATACCTGATGACGGT 180
DB 121 CCACCCCAACACACTTTCCTACTACCAAGTTCCTCAAGATTAGATACCTGATGACGGT 180
QY 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGGGAAACCCAGAAATCTACATTCAA 240
DB 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGGGAAACCCAGAAATCTACATTCAA 240
QY 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTAACTTAACC 300
DB 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTAACTTAACC 300
QY 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGGAGGATAGAGTAAT 360
DB 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGGAGGATAGAGTAAT 360
QY 361 TGGGTGCTATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
DB 361 TGGGTGCTATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
QY 421 ACTGATGCCCCAATACCATTACCCAGTAAAGTTTCAACCTAACAGACTTCTATCTAACA 480
DB 421 ACTGATGCCCCAATACCATTACCCAGTAAAGTTTCAACCTAACAGACTTCTATCTAACA 480
QY 481 ATCTCTCTATAACTGAGCCCAAGAACCGCTGCGCAATTAATCTTGGCAATAGAACTCTGG 540
DB 481 ATCTCTCTATAACTGAGCCCAAGAACCGCTGCGCAATTAATCTTGGCAATAGAACTCTGG 540
QY 541 TTAACGAGAGAGCTTGGAGAACACAGGATTAACAGCGATGACAGAAAGTAAGTA 600
DB 541 TTAACGAGAGAGCTTGGAGAACACAGGATTAACAGCGATGACAGAAAGTAAGTA 600
QY 601 TGGATTTACTATGACGATTAACACCGCTGGCTCCCAAGTTAAGGAGATTGTAGTCCCA 660
DB 601 TGGATTTACTATGACGATTAACACCGCTGGCTCCCAAGTTAAGGAGATTGTAGTCCCA 660
QY 661 ATATAGTTACCGAAACACAGTAATGCTATCTTGAAGTATGGAAGGCAACATTTGGT 720
DB 661 ATATAGTTACCGAAACACAGTAATGCTATCTTGAAGTATGGAAGGCAACATTTGGT 720
QY 721 TGGGAGTATGTTGCAATTTAGAAATAGAACCCCAATCAAGAGGGAACAGTGCAATTTCCA 780
DB 721 TGGGAGTATGTTGCAATTTAGAAATAGAACCCCAATCAAGAGGGAACAGTGCAATTTCCA 780
QY 781 TACGAGCATTTATAGTGTGAGGCAACATTTCAAGCTTACCAATACAGAACTT 840
DB 781 TACGAGCATTTATAGTGTGAGGCAACATTTCAAGCTTACCAATTTACAGAACTT 840

```

```

RESULT 2
US-10-093-037-60
; Sequence 60, Application US/10093037
; Publication No. US2003078397A1
; GENERAL INFORMATION:
; APPLICANT: Jay M. Short
; APPLICANT: Bylina, Edward
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Lam, David E.
; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: 09010-024006
; CURRENT APPLICATION NUMBER: US/10/093, 037
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 09/910, 579
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 09/134, 078
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 08/949, 026
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: US 60/056, 916
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-10-093-037-60

```

```

Query Match      100.0%; Score 960; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 3.4e-272;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAAGTTGGTATCTATCTTAAACAATCTTTTATAGCAGCAATA 60
DB 1 ATGAGCAAGAAAAAGTTGGTATCTATCTTAAACAATCTTTTATAGCAGCAATA 60
QY 61 TATTTTGTAGAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
DB 61 TATTTTGTAGAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
QY 121 CCACCCCAACACACTTTCCTACTACCAAGTTCCTCAAGATTAGATACCTGATGACGGT 180
DB 121 CCACCCCAACACACTTTCCTACTACCAAGTTCCTCAAGATTAGATACCTGATGACGGT 180
QY 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGGGAAACCCAGAAATCTACATTCAA 240
DB 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGGGAAACCCAGAAATCTACATTCAA 240
QY 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTAACTTAACC 300
DB 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTAACTTAACC 300
QY 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGGAGGATAGAGTAAT 360
DB 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGGAGGATAGAGTAAT 360
QY 361 TGGGTGCTATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
DB 361 TGGGTGCTATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
QY 421 ACTGATGCCCCAATACCATTACCCAGTAAAGTTTCAACCTAACAGACTTCTATCTAACA 480

```



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 978
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-228-063-58

Query Match      55.9%; Score 536.8; DB 15; Length 978;
Best Local Similarity 74.4%; Pred. No. 2.3e-147;
Matches 676; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 52 CAGGCAATATATTTGTAGAAAGATCATACCTCTGAGGACAACTCAACTCAATACC 111
Db 70 CATCAATCTACTTGTGAGAGTACCACACTCCGAGGACAACTCCACCTCCAACACC 129
QY 112 TCATCTACACACCCCAACACACTTTCCACTACCAAGTTTCTCAAGATTAGATACCT 171
Db 130 TCCTCACCCTCGCGCAGACACCTCTCCACCAACAGGTGCTCAAGTCCGCTACCG 189
QY 172 GATGACGGTGAAGTGGCCAGGAGCTCTATGTATAGATGGTGGAAACCCAGATTC 231
Db 190 GACGACGGTGAAGTGGCCGCGCCCGGATCGACAAAGGACGGCGCAACCCGAGTTC 249
QY 232 TACATTGAAATAAACCCTATGGAACATCTTAAATGCTACTGGATTGCTGAGATGACGTAC 291
Db 250 TACATCGAGATCAACTCTGGACATCTCAACGACACCGGCTTCGCGAGATGACCTAC 309
QY 292 AATTAAACAGCGGGCTCTTCACTAGCTGCAACAACTTGACAACTTTCTTGGAGGAT 351
Db 310 AACCTCACTAGTGGGGTGTCTCACTACGTGACGACGTGCAACATCGTGTCTCGGAC 369
QY 352 AGAAGTAAATGGTGCATGATACCCCGAATATTTCTATGGAACAAGCCATGAAATGCA 411
Db 370 CGCTCAACTGGGTGACGGCTTACCCGGAATCTTCTACGGCAACAAGCCGTGGAACGCC 429
QY 412 AACTACGCAACTGATGGCCCAATACCATTAACCAAGTAAAGTTTCAAACTAACAGACTTC 471
Db 430 AACTACGCAACGACGGCCGATCCGCTCCGCTCCAAAGGTGTCCAACTCACCGACTTC 489
QY 472 TATCTAACAATCTCTATAAATTTAGCGCCAGAACGCGCTGCCAATTAATCTTGGCAATA 531
Db 490 TACCTCAACATCTCTACAAGCTGACGGCGAAGAACGGTCTCCGGATCAACTTCGCCATC 549
QY 532 GAATCTCTGGTTAACGAGAAAGCTTGGAGAAACACAGGAAATTAACAGCGATGAGCAAGAA 591
Db 550 GAGTCTCTGCTCACCGCGAGGCTGGCGCACCCAGCCGATCACTCCGACGACGAGAG 609
QY 592 GTAATGATATGGATTACTATGACGGATTAACAACGGCTGGCTCCAAAGTTAAGGAGATT 651
Db 610 GTGATGATCTGATCTACTACGACGGCTCCAGCCCGCGGCTCCAAAGGTGAAGAGATC 669
QY 652 GTAGTCCCAATAATATTAACGGAACACAGTAATGCTACATTTGAAGTATGAAGGCA 711
Db 670 GTGGTCCGATCATGCTGAAACGACCCCGGTGACGCCACTTCGAGGTGGAAGGCC 729
QY 712 AACATTGGTGGAGTATGTTGCAATTTAGATAATAAGACCCCAATCAAGAGGGGAACAGTG 771
Db 730 AACATCGGTGGAGTACGTGGCTTCGCGATCAAGACCCCGATCAAGAGGGGACCGGTG 789
QY 772 ACAATTTCATAGGAGATTTATAGTGTTCGACGCAACATTTCAAGCTTACCAATAC 831
Db 790 ACCATCCGTTAGGGCGCTTCACTCCGTCGCGCCCAACATCTCTCTCCCGGAATAC 849
QY 832 ACAGAACTTTACTTACGACGCTGGAGATTGGAATTTGGAATTTGGAACGCCAACGACTACC 891
Db 850 ACCGAGAAAGTACCTCGAGGACGTGGAGATGGCACCGGAGTTCCGACACCCCGTCCACCA 909
QY 892 TCCGCCACCTAGAGTGGTGGATCAACAACATACTAATCTCTTCTAGATAGACCTTT 951
Db 910 TCCGCCACCTCGAGTGGTGGATCAACAACATCACTACCCCTCACCCGCTCGACCGCCGCTC 969
```

```
QY 952 ATTTCTCTA 959
Db 970 ATCTCTA 977

RESULT 5
US-10-228-063-56
; Sequence 56, Application US/10228063
; Publication No. US20030135885A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; TITLE OF INVENTION: Self-processing Plants and Plant Parts
; FILE REFERENCE: 109846 317
; CURRENT APPLICATION NUMBER: US/10/228, 063
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
; US-10-228-063-56

Query Match      55.6%; Score 534; DB 15; Length 903;
Best Local Similarity 74.5%; Pred. No. 1.5e-146;
Matches 672; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 58 ATATATTTGTAGAAAGTATCATACCTCTGAGGACAACTCAACTCAATACCTCATCT 117
Db 1 ATCTACTTGTGGAGAGTACCACACTCCGAGGACAACTCCACCTCCAACACCTCTCTCC 60
QY 118 ACACCAACCCCAACAACTTTCCACTACCAAGTTTCTCAAGATTAGATACCTCTGATGAC 177
Db 61 ACCCGCGGACGACCACTCTCCACCAACAGGTGCTCAAGTCCGCTACCCGAGCAGC 120
QY 178 GGTGATGCGCAGGAGCTCTATTTGATAAGGATGGTGAATGGGAACCCAGAAATTCACATT 237
Db 121 GGCGAGTGGCGCGCGCCCGATCGACAGGACGGCGACCGGACCCGAGTTCTACATC 180
QY 238 GAATAAACCCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGATCAATTTA 297
Db 181 GAGATCAACCTCTGGAACATCTCTAACGCCACCGGCTTCGCGAGATGACCTTACACCTC 240
QY 298 ACACGCGGCTCTTCACTACCTTCAACAACTTGAACAANTTGTCTTGAGGATAGAAGT 357
Db 241 ACTAGTGGGTGTCTCACTACGTGACGAGCTCGACCAACATCGTCTCGCGACCGCTCC 300
QY 358 AATTGGTGCATGGATACCCCGAATATTTCTATGGAACAAGCCATGGAATGCAACTAC 417
Db 301 AACTGGTGCATGGTACCCCGAATCTTCTACGGCAACAGCCGTGGAACCCCACTAC 360
QY 418 GCAACTGATGGCCCAATACCATTAACCAAGTTTCAAACTAACAGACTTCTATCTA 477
Db 361 GCCACGACGCGCCGATCCCGCTCCCGTCCAAAGGTGCCAATCTCACCGACTTCTACCTC 420
QY 478 ACATCTCTATAAATTGAGCCCAAGAACGCGCTTGCATTAATTAATCTTGCATAGAAATCC 537
Db 421 ACCATCTCTACAAGCTCGAGCCGAGAACGCTTCCCGATCAACTTCGCCATCCAGTCC 480
QY 538 TGGTTAACGAGAGAGCTTGGAGAAACACAGGAAATTAACAGCGATGAGCAAGAGTAAATG 597
Db 481 TGGCTCACCGCGAGGCTGGCGCACCCAGCCATCACTCCGACGACGAGGATGATG 540
QY 598 ATATGATTTACTATGACGGATTAACAACGGGTGCTCCAAAGTTTAAAGGATTTGATGCT 657
Db 541 ATCTGGATCTACTACGACGGCTCCAGCCCGCGGCTCCAAAGGTGAGGAGATCTGTGGT 600
QY 658 CCAATAATAGTTAAACGGAACACAGTAAATGCTACATTTGAAGTATGAAGCAAACTT 717
Db 601 CGATCATCTGTAAGCGGACCCCGGCTGAACGCCACTTCGAGGTGTGAAGCCCAACTC 660
QY 718 GGTGGAGTATGTTGCAATTTAGAAATAAGACCCCAATCAAGAGGGAACAGTACAATT 777
```



Db 661 GGCTGGAGTACGTGGCTTCCGCATCAAGACCCCGATCAAGGAGGCGACCGTGAACATC 720  
 QY 778 CCATACGAGCATTTATAGTGTTCGAGCCACATTTCAAGCTTACCAATTTACACAGAA 837  
 Db 721 CCGTACGGCGCTTCACTCGTGGCCGCCAACATCTCTCCCTCCGAACTACACCGAG 780  
 QY 838 CTTTACTTTAGGAGCGTGGAGATGGAACTGAGTTTGGAAAGCGCCAAAGCACTACCTCCGCC 897  
 Db 781 AAGTACCTCGAGGAGCGTGGAGATCGGACCGAGTGGCGACCCCGTCCACCACTCCGCC 840  
 QY 898 CACCTAGAGTGGTGCATACAAATACATACATCACTCTCTAGATAGACCTCTTATTCC 957  
 Db 841 CACCTCGAGTGGTGCATACCAATACATCACTCCCTCACCCTCGACCCCGCTCATCTCC 900  
 QY 958 TA 959  
 Db 901 TA 902

RESULT 6  
 US-10-369-493-26728  
 ; Sequence 26728, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 26728  
 ; LENGTH: 825  
 ; TYPE: DNA  
 ; ORGANISM: Thermotoga maritima  
 US-10-369-493-26728

Query Match 8.8%; Score 84.2; DB 16; Length 825;  
 Best Local Similarity 50.2%; Pred. No. 6, 7e-14;  
 Matches 305; Conservative 0; Mismatches 278; Indels 24; Gaps 3;

Db 557 AGTGGGATGTTTACTTTGCAACCGTGGGATGGGATTAACCTTCTTTCAGACTGACAACAC 616  
 QY 752 CAATCAAGAGGGAACAGTGAACAATTCATACGGAGCATTTATAAGTGTTCAGAGCCACA 811  
 Db 617 CGATGAAGAGGGAAGGTGAAATCAACGTGAAGCACTTCGTTAGAAAGCCGCGGAG 676  
 QY 812 TT-----TCAGCTTACCAANTTACACAGAACTTTTACTTTAGAGGAGCTGG 856  
 Db 677 TTGTCGAAGAGCACTCAACGAGATAGACAATTTGAGAGAGCTGTATTTCTCGCTGG 736  
 QY 857 AGATTCGAAGTGGTGGAAAGCGCAAGCACTACCTCCGCCACCTTAGAGTGTGTGATCA 916  
 Db 737 AGATCGAAGCGAGTTTGGAGATCGGACACACACGCGGGAATTCGCTGACTTTCA 796  
 QY 917 CRAACAT 923  
 Db 797 GGGACTT 803

RESULT 7  
 US-10-027-632-197533/c  
 ; Sequence 197533, Application US/10027632  
 ; Publication No. US20020198371A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 197533  
 ; LENGTH: 585  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-197533

Query Match 4.1%; Score 39; DB 13; Length 585;  
 Best Local Similarity 50.3%; Pred. No. 1, 2;  
 Matches 93; Conservative 1; Mismatches 91; Indels 0; Gaps 0;

QY 638 AAGTTAAGGAGATGTAGTCCCAATATAGTTAAACGGAACACACAGTAAATGCTTACATTG 697  
 Db 523 AAGTACACAGATATGCTAAGAAAGGAGAGAAATGGAATCATATAAATGCTAAATTA 464  
 QY 698 AAGTATGGAAGCAACATGTTGGAGATGTTGTCATTTAGAAATAAGAGCCCCAATCA 757  
 Db 463 AAGTACAAAGGAGAGAAAGAGAGAGACAAAATAGTAAAAAGACAGAGTAACA 404  
 QY 758 AAGAGGGAACAGTGAACAATTCATACGAGCATTTTATAGTGTTCAGAGCAACATTTCAA 817  
 Db 403 AAGAGAAACAGTAAACAATGCTGTTGATATTATTCCTCACTATATTAGTAATCAC 344  
 QY 818 GCTTA 822  
 Db 343 AAGTA 339

```

RESULT 8
US-10-027-632-197533/c
; Sequence 197533, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 197533
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-197533

```

```

Query Match 4.1%; Score 39; DB 16; Length 585;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 93; Conservative 1; Mismatches 91; Indels 0; Gaps 0;

QY 638 AAGTTAAGGAGATGTTAGTCCCAATTAAGTTAAGCGAACACAGTAAATGCTACATTTG 697
Db 523 AAGTACAACAGATATGCTTAAGAAAGGAGAGAAATGGAATCATATAAATGCTAAATTAA 464
QY 698 AAGTATGGAAGCGAACATGTTGTTGGAGTATGTTGCTATTTAGATAAAGACCCCAATCA 757
Db 463 AACTACAAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
QY 758 AAGAGGGAACAGTGAATTTCCATACGAGGAGCATTTATAGTGTTCGACGCCAATTTCAA 817
Db 403 AAGAGAAACAGTGAATTTCCATACGAGGAGCATTTATAGTGTTCGACGCCAATTTCAA 344
QY 818 GCTTA 822
Db 343 AAGTA 339

```

```

RESULT 9
US-08-781-986A-611/c
; Sequence 611, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

```

```

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 611:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-611

Query Match 3.9%; Score 37.8; DB 8; Length 728;
Best Local Similarity 45.5%; Pred. No. 3;
Matches 135; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATCAGCAAGAAATGATA 600
Db 456 TTAATACCAATGAATGATAATCTACTTACTAGTAGCAATGTAATGATAGTAA 397
QY 601 TGGATTACTATGACGGATTACAAACCGCTGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660
Db 396 ATACTTCTCTGTAGTATTTGCTTCAATTCGATGTAAATTAATTCATGTAATCCAA 337
QY 661 ATATAGTTAAACGAGACACAGTAATGCTACATTTGAAAGTATGGAAGCAACATTTGGT 720
Db 336 GCTACAGATAAAGGAAAAATACCAATCCACTTTGCAATGGAGATAATCTAGTTTCT 277
QY 721 TGGGAGTATGTTCCATTTAGATTAAGACCCCAATCAAGAGGAGACAGTCAATTTCA 780
Db 276 TGAAGGTAGCTGGAATATATAATTTGAATCCATTTAAACAATAGCTTTAGAGGTCA 217
QY 781 TACGAGCATTTTAAAGTGTTCGAGCCCAATTTTCAAGCTTTACCAATTTACACAGAA 837
Db 216 GTTATAAAACTAAAGTAATCGAATCGAATTAATTTCAACACAGGTAATAATGTTGAA 160

```

```

RESULT 10
US-10-329-624-611/c
; Sequence 611, Application US/10329624
; Publication No. US20040043037A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text

```

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/329,624  
 FILING DATE: 27-Dec-2002  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/956,171  
 FILING DATE: October 20, 1997  
 APPLICATION NUMBER: 60/009,861  
 FILING DATE: January 5, 1996  
 APPLICATION NUMBER: 08/781,986  
 FILING DATE: January 3, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mark J. Hyman  
 REGISTRATION NUMBER: 46,789  
 REFERENCE/DOCKET NUMBER: PB248P1D1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (240) 314-1224  
 TELEFAX: (301) 309-8439  
 INFORMATION FOR SEQ ID NO: 611:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 728 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 611:  
 US-10-329-624-611

Query Match 3.9%; Score 37.8; DB 13; Length 728;  
 Best Local Similarity 45.5%; Pred. No. 3;  
 Matches 135; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 541 TTAACGAGAGAGCTTGAGAACACAGCAATTAACGCGATGAGCAAGTAATGATA 600  
 Db 456 TTAATACCAATGAATGATATATCTACTTACTAGTCAATGTAATGATAGTAA 397  
 QY 601 TGGATTACTATGACGGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGATCCCA 660  
 Db 396 ATAACCTTCCTGTAGTTTGGCTTCAATTCGATGTAATTAATAATCAATGTAATCCAA 337  
 QY 661 ATAATAGTTACGAGACACAGTAATGCTACATTTGAGTATGAGGCGCAACATTTGGT 720  
 Db 336 GCTACAGATAAGGAATAATCAACATCCATTTGCAATGAGATATCTTGTGTCT 277  
 QY 721 TGGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGGAACAGTACAAATCCA 780  
 Db 276 TGAAGTAGACTGGAATATATAATTAATGATCCCAATTAACAAATAGCTGTAGAGGTCA 217  
 QY 781 TACGAGCATTTAATGATGTTGCGCCCAACATTTCAAGCTTACCAATTTACACAGAA 837  
 Db 216 GTTATAAACTAAAGTAATCGAAGCATTAATAATCAACACAGGTAATAAATGGTGAA 160

RESULT 11  
 US-09-917-800A-1558/c  
 Sequence 1558, Application US/09917800A  
 Patent No. US20020119462A1  
 GENERAL INFORMATION:  
 APPLICANT: Mendrick, Donna  
 APPLICANT: Porter, Mark  
 APPLICANT: Johnson, Kory  
 APPLICANT: Castle, Arthur  
 APPLICANT: Elashoff, Michael  
 APPLICANT: Gene Logic, Inc.  
 TITLE OF INVENTION: Molecular Toxicology Modeling  
 FILE REFERENCE: 44921-5038-US  
 CURRENT APPLICATION NUMBER: US/09/917,800A  
 CURRENT FILING DATE: 2001-07-31  
 PRIOR APPLICATION NUMBER: US 60/222,040  
 PRIOR FILING DATE: 2000-07-31  
 PRIOR APPLICATION NUMBER: US 60/222,880  
 PRIOR FILING DATE: 2000-11-02  
 PRIOR APPLICATION NUMBER: US 60/290,029  
 PRIOR FILING DATE: 2001-05-11  
 PRIOR APPLICATION NUMBER: US 60/290,645

PRIOR FILING DATE: 2001-05-15  
 PRIOR APPLICATION NUMBER: US 60/292,336  
 PRIOR FILING DATE: 2001-05-22  
 PRIOR APPLICATION NUMBER: US 60/295,798  
 PRIOR FILING DATE: 2001-06-06  
 PRIOR APPLICATION NUMBER: US 60/297,457  
 PRIOR FILING DATE: 2001-06-13  
 PRIOR APPLICATION NUMBER: US 60/298,884  
 PRIOR FILING DATE: 2001-06-19  
 PRIOR APPLICATION NUMBER: US 60/303,459  
 PRIOR FILING DATE: 2001-07-09  
 NUMBER OF SEQ ID NOS: 1740  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1558  
 LENGTH: 2338  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 FEATURE:  
 OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM\_012674  
 US-09-917-800A-1558

Query Match 3.9%; Score 37.2; DB 9; Length 2338;  
 Best Local Similarity 49.0%; Pred. No. 9.1;  
 Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 550 GRAGCTTGGAGAACACAGCAATTAACGCGATGAGCAAGTAATGATATGGATTAC 609  
 Db 1997 GAATGTGGGAGGTCACAGAGTACCAACCCCTTAACAGGAGTCTTAAAGTCAATTTTC 1938  
 QY 610 TATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCAATAATAGTT 669  
 Db 1937 TTTAGGGGTATGGCTTTTCATATATTGGCAGTAACAGTGCATGGTGTACACTCATGAGT 1878  
 QY 670 AACGGAACACAGTAATGCTACATTTGAGTATGAGGCGCAACATTTGGTGGGAGTAT 729  
 Db 1877 ATATGACGACAGTAATTTGAGTCTTTTGGCTTATTAATAAATACATGAAGTTAGGAAGT 1818  
 QY 730 GTTGCAATTTAGATAAAGACCC 751  
 Db 1817 GAACTTTAGAAACGGGACTC 1796

RESULT 12  
 US-10-142-426-144  
 Sequence 144, Application US/10142426  
 Publication No. US20040048333A1  
 GENERAL INFORMATION:  
 APPLICANT: Baker, Kevin P.  
 APPLICANT: Beresini, Maureen  
 APPLICANT: DeForge, Laura  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Filvaroff, Ellen  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Gerritsen, Mary E.  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Godowski, Paul J.  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Sherwood, Steven  
 APPLICANT: Smith, Victoria  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tamas, Daniel  
 APPLICANT: Watanabe, Colin K  
 APPLICANT: Wood, William  
 APPLICANT: Zhang, Zemin  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 FILE REFERENCE: P330RJC224  
 CURRENT APPLICATION NUMBER: US/10/142,426  
 CURRENT FILING DATE: 2002-05-09  
 Prior Application removed - See File Wrapper or Palm  
 NUMBER OF SEQ ID NOS: 550  
 SEQ ID NO 144  
 LENGTH: 520

```
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-142-426-144

Query Match      3.8%; Score 36.8; DB 13; Length 520;
Best Local Similarity 10.2%; Pred.No.4.9; Mismatches 0; Gaps 0;
Matches 53; Conservative 143; Indels 322;

QY   225 AGAATTCTACATGAATAAACTTGGAACACTTTCTTAATGCTACTGAGATTGTGCAGAT 284
       ::::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    1 MVARVGLLRALQLLWGHLDQAERGGQELRKEAEAFLEKYGYLNEQVPKAPTSTFRS 60

QY   285 GAGGTACAATTTAAACAGCGCGTCCTTCACTACGTCGCAACCACTTGACACATTCGCTT 344
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    61 DAIRAFOWSQLPVSGVLDRATLRMTPRCGVTDTNSYAABERIIDLFAHRTKWRK 120

QY   345 GAGGGATAGAAGTAATTGGTGCATGCCCATACCCTCCGAATATTTCTATGAAAACAGCCTG 404
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    121 KRFAPQGNKWTKQHLSRLVNWPHELPEPAVRGVRAAFOLWNVSALFEWEAPATGPAD 180

QY   405 GAATGCAAACTACGCAACTGATGCGCCAATACCATTACCAGTAAAGTTTCAAACCTAAC 464
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    181 IRLTFQGDNHLGNFDPGGALAHAFLLRRGEAHFDQDERWSLRRGRNLFVVLAH 240

QY   465 AGACTTCTATCTAACAATCTCTATAAATGAGCCCCAAGAGCGCTGCCAATTAACCTT 524
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    241 EIGHTLGLTHSPAPRALMAPYYKLGRDALLSDVDLVAVQSLYKGPLGGSVAVOLPKLP 300

QY   525 CGCAATAGAACTCCTGGTTAAACAGAGAGAGCTTTGGAGAACACACAGGAATTAACAGCGATGA 584
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    301 TDFETWDSSYPQRRPETQGPKYCHSSFDAITVDRQQQLYIFKGSHEFWAADVSEPR 360

QY   585 GCAAGAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 644
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    361 FLQERWVGLPNIIEAAVSLNDGDFFFKGRCWRFRGPKVPWGVLPOLCRAGGLPHPD 420

QY   645 GGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 704
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    421 ALFPFLRLILFKARYVYLARGQLVEPYPRSLQDWGIIPEEVS GALPRDGSIIF 480

QY   705 GAAGCAAACTGTTGGAGTAGTTGTCATTAGAA 742
       ::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    481 RDRYWRLDQAKLOATTSGRWATELPMWGCHANS SSA 518


RESULT 13
US-10-123-155-144
; Sequence 144, Application US/10123155
; Publication No. US20030068794AI
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: Deforge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wel-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC30
; CURRENT APPLICATION NUMBER: US/10/123,155
; PRIOR FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
```

; CURRENT APPLICATION NUMBER: US/10/146,731  
; CURRENT FILING DATE: 2002-05-15  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 144  
; LENGTH: 520  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-146-731-144

Query Match 3.8%; Score 36.8; DB 15; Length 520;  
Best Local Similarity 10.2%; Pred. No. 4.9; Indels 0; Gaps 0;  
Matches 53; Conservative 143; Mismatches 322; Indels 0; Gaps 0;  
QY 225 AGAATTCACATTAACCAATGAACATCTTAATCTACTGATTCGTGAGAT 284  
DB 1 MVARVGLLLRALQLLLWGHLDQAQPAERGGQBLRKEAEAFLEKYLYNEQVPKAPTSTRFS 60  
QY 285 GACGTACAATTTAACCGCGGCTTCCTCACTAGTCCCAACATTCGACACATTTGCTT 344  
DB 61 DAIRAFOWVSQLPVSGVLDRAITLRQMTPRCGVTDITNSYAAWAERISDLFARHETKMRK 120  
QY 345 GAGGATAGAAGTAATTCGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATG 404  
DB 121 KRPAKQGNKWKYKQHLVYLVNWPHELPEPAVRGAVRAAFQWLSNVSALEFWEAATGPAD 180  
QY 405 GAATGCAAACTACGCACTGATGCGCCCAATACCATTACCCAGTAAAGTTTCAAACTAAC 464  
DB 181 IRLTFFQGDHNDGLGNAPDGGALAHAFPRRGEAHFPDQERWSLSRRRGRNLFVVLAH 240  
QY 465 AGACTTCTATCAAACTCTCTATAAATTTAGCCCAAGACCGGCTGCCAATTAATT 524  
DB 241 EIGHTLGLTHSPAPALMAPYKRLGRDALLSWDDVLAVQSLYKPLGGSVAVQLPGKLF 300  
QY 525 CGCAATAAGTCTGTTTAAACGAGAGAGCTTGAGACCAACAGGAATTAACAGCAATGA 584  
DB 301 TDFTWDSYSPQRRPETQPKYCHSSPDALTVDROQQLYIFKGSHEFWAADGNVSEPR 360  
QY 585 GCAAGAAGTAATGATGATGATTTACTATGACGATTACACCGGCTGCCCAAGTTAA 644  
DB 361 PLQERWVGLPPIEAAAASVINDGDFYFFKGRGCRWRFPKPVWGLPQLCRAGGLPRHPDA 420  
QY 645 GGAGATTGTAGTCCCAATATAGTTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATG 704  
DB 421 ALFFPPLRLILFKGARYYVLARGQLQVEPYPRSLQDWGGIPEEVSGALPRPDGSIIF 480  
QY 705 GAAGCAAAACATTTGGTGGAGTATGTTGCAATTTAGAA 742  
DB 481 RDRYWRLDQAKLQATTSGRWATELPMWGCWHANSOSA 518

## RESULT 15

US-10-140-472-144  
; Sequence 144, Application US/10140472  
; Publication No. US2003013888A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C168  
; CURRENT APPLICATION NUMBER: US/10/140,472  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 144  
; LENGTH: 520  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-140-472-144

Query Match 3.8%; Score 36.8; DB 15; Length 520;  
Best Local Similarity 10.2%; Pred. No. 4.9; Indels 0; Gaps 0;  
Matches 53; Conservative 143; Mismatches 322; Indels 0; Gaps 0;  
QY 225 AGAATTCACATTAACCAATGAACATCTTAATCTACTGATTCGTGAGAT 284  
DB 1 MVARVGLLLRALQLLLWGHLDQAQPAERGGQBLRKEAEAFLEKYLYNEQVPKAPTSTRFS 60  
QY 285 GACGTACAATTTAACCGCGGCTTCCTCACTAGTCCCAACATTCGACACATTTGCTT 344  
DB 61 DAIRAFOWVSQLPVSGVLDRAITLRQMTPRCGVTDITNSYAAWAERISDLFARHETKMRK 120  
QY 345 GAGGATAGAAGTAATTCGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATG 404  
DB 121 KRPAKQGNKWKYKQHLVYLVNWPHELPEPAVRGAVRAAFQWLSNVSALEFWEAATGPAD 180  
QY 405 GAATGCAAACTACGCACTGATGCGCCCAATACCATTACCCAGTAAAGTTTCAAACTAAC 464  
DB 181 IRLTFFQGDHNDGLGNAPDGGALAHAFPRRGEAHFPDQERWSLSRRRGRNLFVVLAH 240  
QY 465 AGACTTCTATCAAACTCTCTATAAATTTAGCCCAAGACCGGCTGCCAATTAATT 524  
DB 241 EIGHTLGLTHSPAPALMAPYKRLGRDALLSWDDVLAVQSLYKPLGGSVAVQLPGKLF 300  
QY 525 CGCAATAAGTCTGTTTAAACGAGAGAGCTTGAGACCAACAGGAATTAACAGCAATGA 584  
DB 301 TDFTWDSYSPQRRPETQPKYCHSSPDALTVDROQQLYIFKGSHEFWAADGNVSEPR 360  
QY 585 GCAAGAAGTAATGATGATGATTTACTATGACGATTACACCGGCTGCCCAAGTTAA 644  
DB 361 PLQERWVGLPPIEAAAASVINDGDFYFFKGRGCRWRFPKPVWGLPQLCRAGGLPRHPDA 420  
QY 645 GGAGATTGTAGTCCCAATATAGTTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATG 704  
DB 421 ALFFPPLRLILFKGARYYVLARGQLQVEPYPRSLQDWGGIPEEVSGALPRPDGSIIF 480  
QY 705 GAAGCAAAACATTTGGTGGAGTATGTTGCAATTTAGAA 742  
DB 481 RDRYWRLDQAKLQATTSGRWATELPMWGCWHANSOSA 518

Search completed: July 2, 2004, 02:27:28  
Job time : 653 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 21:58:24 ; Search time 102 Seconds  
(without alignments)  
5223.067 Million cell updates/sec

Title: US-09-914-543-45

Perfect score: 960

Sequence: 1 atgagcaagaaagtctgt.....atagacctttattcttaa 960

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	960	4	US-09-134-078-60
2	114.2	11.9	867	3	US-08-995-280C-1
3	53.8	5.6	7218	1	US-08-232-463-14
4	37.8	3.9	728	4	US-08-956-171E-611
5	36	3.8	936	4	US-09-636-382A-16
6	34.8	3.6	1242	4	US-09-134-000C-3076
7	34.8	3.6	3146	4	US-09-620-312D-277
8	34.8	3.6	15016	4	US-09-601-198-60
9	34.2	3.6	1161	4	US-09-328-352-1845
10	33.8	3.5	1664976	4	US-08-916-421B-1
11	33.4	3.5	3543	3	US-09-224-024-27
12	33.4	3.5	3543	5	PCT-US94-07802-27
13	33.2	3.5	612	3	US-09-385-982-400
14	32.8	3.4	832	4	US-09-621-976-2813
15	32.8	3.4	1422	1	US-08-319-704-5
16	32.4	3.4	3735	4	US-09-543-681A-2102
17	32.4	3.4	90541	4	US-09-759-359A-3
18	32.2	3.4	645	4	US-09-328-352-661
19	32	3.3	1652	4	US-09-375-140-8
20	32	3.3	11049	4	US-10-204-708-22
21	31.8	3.3	828	4	US-09-371-056-9
22	31.8	3.3	1664	4	US-09-397-787-68
23	31.8	3.3	1664	4	US-09-397-787-267
24	31.8	3.3	1892	4	US-09-276-438-12
25	31.8	3.3	4242	4	US-09-276-438-11
26	31.8	3.3	4487	4	US-09-484-970B-56
27	31.8	3.3	44453	4	US-09-146-053-5

28 31.6 3.3 480 4 US-09-621-976-8465 Sequence 8465, Appl  
29 31.6 3.3 1185 3 US-08-975-762-15 Sequence 15, Appl  
30 31.6 3.3 1185 3 US-08-821-324-15 Sequence 15, Appl  
31 31.6 3.3 1185 3 US-09-295-028-15 Sequence 15, Appl  
32 31.6 3.3 1185 4 US-09-106-582-15 Sequence 15, Appl  
33 31.6 3.3 1185 4 US-09-159-469-15 Sequence 15, Appl  
34 31.6 3.3 1185 4 US-09-693-542-15 Sequence 15, Appl  
35 31.6 3.3 2129 3 US-08-975-762-39 Sequence 39, Appl  
36 31.6 3.3 2129 3 US-09-295-028-39 Sequence 39, Appl  
37 31.6 3.3 2129 4 US-09-106-582-39 Sequence 39, Appl  
38 31.6 3.3 2129 4 US-09-159-469-39 Sequence 39, Appl  
39 31.6 3.3 2129 4 US-09-693-542-39 Sequence 39, Appl  
40 31.6 3.3 3998 3 US-09-066-046-5 Sequence 5, Appl  
41 31.6 3.3 6376 4 US-09-405-728-1 Sequence 1, Appl  
42 31.6 3.3 99500 4 US-09-798-096-10 Sequence 10, Appl  
43 31.4 3.3 261 3 US-08-906-769-160 Sequence 160, App  
44 31.4 3.3 261 3 US-08-906-769-160 Sequence 160, App  
45 31.4 3.3 261 3 US-08-639-075A-160 Sequence 160, App

#### ALIGNMENTS

RESULT 1  
US-09-134-078-60  
; Sequence 60, Application US/09134078  
; Patent No. 6368844

#### GENERAL INFORMATION:

APPLICANT: Bylina, Edward J.  
TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92121

#### COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/134,078  
FILING DATE: 13-AUG-1998  
CLASSIFICATION: 435

#### PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/949,026  
FILING DATE: 10-OCT-1997  
APPLICATION NUMBER: 60/056,916  
FILING DATE: 06-DEC-1996

#### ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347

#### REFERENCE/DOCKET NUMBER:

09010/024002

#### TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465

#### INFORMATION FOR SEQ ID NO:

60:

#### SEQUENCE CHARACTERISTICS:

LENGTH: 960 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...957

US-09-134-078-60

Query Match 100.0%; Score 960; DB 4; Length 960;  
Best Local Similarity 100.0%; Pred. NO. 1.2e-291;



Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ATGAGCAAGAAAAGTTGCTCATCTATCTATCTTAACAATCCCTTTTAGTACAGGCAATA	60
Db	1	ATGAGCAAGAAAAGTTGCTCATCTATCTATCTTAACAATCCCTTTTAGTACAGGCAATA	60
Qy	61	TATTTTGTAGAAAAGTATCATACCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA	120
Db	61	TATTTTGTAGAAAAGTATCATACCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA	120
Qy	121	CCACCCCAACCAACTTTCCTACTACCAAGGTTCTCAAGATTAGATACCTCGATGACGGT	180
Db	121	CCACCCCAACCAACTTTCCTACTACCAAGGTTCTCAAGATTAGATACCTCGATGACGGT	180
Qy	181	GAGTGCCAGAGCTCCATTGATTAAGATGGTGTGGGAACCCGAGAATTTCTACATTGAA	240
Db	181	GAGTGCCAGAGCTCCATTGATTAAGATGGTGTGGGAACCCGAGAATTTCTACATTGAA	240
Qy	241	ATAAACCTATGGAACATTTCTAATGCTACTGGAATTTGATGACGCTCAATTTAAC	300
Db	241	ATAAACCTATGGAACATTTCTAATGCTACTGGAATTTGATGACGCTCAATTTAAC	300
Qy	301	AGCGGCTCTCTCACTAGCTCAACCAACTTGCAACATTTGCTTGGGGATAGAGTAAT	360
Db	301	AGCGGCTCTCTCACTAGCTCAACCAACTTGCAACATTTGCTTGGGGATAGAGTAAT	360
Qy	361	TGGTGTCATGGATACCCGGAATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA	420
Db	361	TGGTGTCATGGATACCCGGAATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA	420
Qy	421	ACTGATGGCCCAATACCACTTACCCAGTAAAGTTTCAAACTTAACAGATTTCTATTAACA	480
Db	421	ACTGATGGCCCAATACCACTTACCCAGTAAAGTTTCAAACTTAACAGATTTCTATTAACA	480
Qy	481	ATCTCTATAAATTGAGCCCAAGAACCGCTTGAAGTATGGAAGCAAGATTAATGATA	540
Db	481	ATCTCTATAAATTGAGCCCAAGAACCGCTTGAAGTATGGAAGCAAGATTAATGATA	540
Qy	541	TTAACGAGAGAGCTTGGAGAACCAACAGGAAATTAACAGGATGACCAAGAGTAATGATA	600
Db	541	TTAACGAGAGAGCTTGGAGAACCAACAGGAAATTAACAGGATGACCAAGAGTAATGATA	600
Qy	601	TGGATTACTATGACGGATTAACAACGGCTGGCTCAAAGTTAAGGAGATTGTAGTCCCA	660
Db	601	TGGATTACTATGACGGATTAACAACGGCTGGCTCAAAGTTAAGGAGATTGTAGTCCCA	660
Qy	661	ATATAGTTTAAACGACACCACTTAATGCTATGTAAGTATGGAAGCAAGATTAATGATA	720
Db	661	ATATAGTTTAAACGACACCACTTAATGCTATGTAAGTATGGAAGCAAGATTAATGATA	720
Qy	721	TGGGAGTATGTTGCAATTTAGAATAAAGACCCCAATCAAGAGGGGAACAGTGAATCCA	780
Db	721	TGGGAGTATGTTGCAATTTAGAATAAAGACCCCAATCAAGAGGGGAACAGTGAATCCA	780
Qy	781	TACGGAGCATTTATAGTCTTCAGCCCAACATTTCAAGTTTACCAAAATACACAGACTT	840
Db	781	TACGGAGCATTTATAGTCTTCAGCCCAACATTTCAAGTTTACCAAAATACACAGACTT	840
Qy	841	TACTTAGAGGACGTGGAGATTGGAACCTGAGTTTGGAAACCCCAAGCACTACCTCCGCCAC	900
Db	841	TACTTAGAGGACGTGGAGATTGGAACCTGAGTTTGGAAACCCCAAGCACTACCTCCGCCAC	900
Qy	901	CTAGAGTGGTGGATCAAAACATTAACATTAACCTCTCTAGATAGACCTCTTATTTCTTAA	960
Db	901	CTAGAGTGGTGGATCAAAACATTAACATTAACCTCTCTAGATAGACCTCTTATTTCTTAA	960

RESULT 2  
US-08-995-280C-1  
; Sequence 1, Application US/08995280C  
; Patent No. 6043075  
; GENERAL INFORMATION:  
; APPLICANT: Bjornvad, Mads Eskelund

APPLICANT: Schulein, Martin  
APPLICANT: No. 6043075revang, Iben Angelica  
TITLE OF INVENTION: A No. 6043075el Endoglucanase  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 6043075o No. 6043075disk of No. 6043075th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/995,280C  
FILING DATE: 19-Dec-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol A.  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 4906.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 867 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-995-280C-1

Query Match 11.9%; Score 114.2; DB 3; Length 867; Best Local Similarity 51.3%; Pred. No. 6.6e-26; Matches 291; Conservative 0; Mismatches 273; Indels 3; Gaps 1;			
Qy	230	TCCTACATGAAATAAACCCTATGGAACATTTCTTAATCTACTGATTTGCTGAGATGACCT	289
Db	137	TCACCTTAGAATCAACTTTTGGAAATTTGCAAACTATGAGGAATACATGATGGCAT	196
Qy	290	ACAAATTTAACCCAGCGGCTCTTCACTACGTCCAACTTGAACCACTTGACAACTTGTGAGG	349
Db	197	TTTATAAGAAAGAGATACCTGTTGAATATATATCCGACATAAAAAACATAGTACTTAAG	256
Qy	350	ATAGAGTAATTCGGTGCATGATACCCGAAATATTTCTATGAAACAAAGCCATGGAATG	409
Db	257	ATAAATTCATGGTACATGGATATCTTGAAGTCTACTATGGGTACAAACCATGGCTG	316
Qy	410	CAAACTACGCAACTGATGGCCCAATACCAATACCCAGTAAAGTTTCAAACTTAACAGCT	469
Db	317	GCCATGGGAATTCATTTGAGAAATTTAGCTCTTCTTAAAAAGGTATCAGAAATTTCCAG	376
Qy	470	TCATCTAACTCTCTCTATAACTTTGAGCCCAAGACCGCTGCCAATTAACCTTCGCA	529
Db	377	TTCTCTCAATCTAAATAACAATATGGTACGAGAGAAATCTTCTTAATTTTCTTA	436
Qy	530	TAGAACTCTGTTAAACGAGAGAGCTTGGAGAAACAAACAGGAATTAACAGCGATGAGCA	589
Db	437	TGGAAACATGATTAACAAGAACCCCTATCAGAAACCG--TTACTTCAGGGGATATAG	493
Qy	590	AAGTAATGATGATTTACTATGACGGATTAACAACGGCTGGCTCCAAAGTTTARGAGA	649
Db	494	AGATGATGGTATGGCTATATGCTAAATAGACTTTCTCTCGAGGGCGAAAGGTAGGAG	553
Qy	650	TTGTAGTCCCAATAATAGTTTAAACGGAACACCACTAAATGCTACATTTGAAGTATGA	709
Db	554	TAAATATACCTATCATCTTAAACGGTAAATCAAAAGACATTAATCTCGGAAGTATATCT	613

QY 710 CAAACATTGGTGGAGTATGTCATTAGATAAAGACCCCAATCAAAGAGGGAACAG 769  
Db 614 CCCTATGAGTGGAGTACTGCTTAAATCAAAGAAAATATCTTCAAGGACAGG 673  
QY 770 TGACAATTCCATACGAGCATTATAA 796  
Db 674 TAAAAATACCAATAAATGAATTTTGA 700

RESULT 3  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, P.  
; APPLICANT: FALKNER, F.G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14  
Query Match 5.6%; Score 53.8; DB 1; Length 7218;  
Best Local Similarity 5.4%; Pred. No. 1.9e-06;  
Matches 22; Conservative 218; Mismatches 165; Indels 0; Gaps 0;

QY 502 AAGACGCGCTGCCAATTAACCTTCGCAATAGAACTCTGTTAAGCAGAGAGCTTGGAGA 561  
Db 1447 AAGAAATTGGTACRR 1388  
QY 562 ACAACAGGAATTAACAGCAGTACGAGAAGTAATGATGATTACTATGACGGATTA 621  
Db 1387 RRR 1328  
QY 622 CAACCGCTGCTCCAAAGTTAAGAGATTGTAGTCCCAATAATAGTTAACGGACACCA 681  
Db 1327 RRR 1268

QY 682 GTAATGCTACATTTGAAGTATGAGGCAACATTTGGAGTATGTTGCAATTAGA 741  
Db 1267 RRR 1208  
QY 742 ATAAGACCCCAATCAAGAGGCAACAGTACAAATTCATACGAGCATTATAAGTGT 801  
Db 1207 RRR 1148  
QY 802 GCAGCCAACTTCAAGCTTACCAATTAACAGAACTTTACTAGAGCAGTGGAGATT 861  
Db 1147 RRR 1088  
QY 862 GGAACGTAGTTGGAACGCAACACTACTCTCGCCCACTAGAG 906  
Db 1087 RRR 1043

RESULT 4  
US-08-956-171E-611/c  
; Sequence 611, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Faanon  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,171E  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Hyman  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB248P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (240) 314-1224  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 611:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 728 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 611:  
US-08-956-171E-611

Query Match 3.9%; Score 37.6; DB 4; Length 728;  
Best Local Similarity 45.5%; Pred. No. 0.061;  
Matches 135; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 541 TTAACAGAGAACTTGGAGAACACAGGAATTACAGCGATGACGAGCAAGTAATGATA 600

Db 456 TTAATACCAAAATGAATGATAATCTACTTACTAGTACCAATGTAAATGATAGTAAGTAA 397  
QY 601 TGGATTACTATGACGGATTACAACCGGTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660  
Db 396 ATAATCTTCCTTGATAGTTTTCCTCAATTCGATGTAATTAATTAATCAATGTAATCAA 337  
QY 661 ATAATAGTTTACGGAAACCAAGTAATGCTACATTTTGAAGTATGGAAGGCAAAATTTGGT 720  
Db 336 GCTACAGATAAAGGAAATAACCAATCCACTTTGCAATGGAGATAATCCTTAGTTGTCT 277  
QY 721 TGGAGTATGTTGCTATTTAGATAAAGACCCCAATCAAGAGGGAACAGTGCACAAATCCA 780  
Db 276 TGAAGTAGACTGGAATATATAAATGATCCCATTAACAATAAGCTGTTAGAGGTCA 217  
QY 781 TAGGAGAGATTTATAAGTTGTCAGCCCAACATTTCAAGCTTACCAAAATACACAGAA 837  
Db 216 GTTATAAAACTAAAGTAATCGAAGGATTAATTCACCCAGGTAATAAATGGTGAA 160

## RESULT 5

US-09-636-382A-16

; Sequence 16, Application US/09636382A

; Patent No. 6514741

; GENERAL INFORMATION:

; APPLICANT: Presnell, Scott R.

; APPLICANT: Taft, David W.

; TITLE OF INVENTION: TRYPTASE-LIKE POLYPEPTIDE ZTRYPI

; FILE REFERENCE: 99-21

; CURRENT APPLICATION NUMBER: US/09/636,382A

; PRIOR FILING DATE: 2000-08-09

; PRIOR APPLICATION NUMBER: US 60/149,563

; PRIOR FILING DATE: 1999-08-18

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 16

; LENGTH: 936

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Degenerate polynucleotide sequence of human ztryp1

; OTHER INFORMATION: (SEQ ID NO:15)

; NAME/KEY: misc\_feature

; LOCATION: (1)...(936)

; OTHER INFORMATION: n = A,T,C or G

US-09-636-382A-16

Query Match 3.8%; Score 36; DB 4; Length 936;  
Best Local Similarity 25.8%; Pred. No. 0.26;  
Matches 66; Conservative 49; Mismatches 141; Indels 0; Gaps 0;  
QY 111 CTCATCTACACCAACCCCAACCACTTTCCACTACCAAGTTCTCAAGATTAGATACCC 170  
Db 245 CNCAYTGYTNCARMGNTTYAARGAYTNWSYTNVWSNGTNGTNGGNGTNCAYC 304  
QY 171 TGATGACGGTAGTGCACCGAGGCTCTATTGATGAAGATGGTGTGGAACCCAGATT 230  
Db 305 AEMGNCNGARAAYSNACNARYTNCNNTNACNMGNATGGTNATHCAYAARGAYTYW 364  
QY 231 CTCATTGAAATAAACCTATGGAACATCTTAACTGCTACTGCTGAGATTGAGATGACGTA 290  
Db 365 SNAAYTNTATGNSCARGAYATHGCTNYTNYTNAARYTNMGAYWSNATHSNATGWSNC 424  
QY 291 CAATTTAACCGCGGCTCTTCACTAGTCCCAACCACTTGCAACACATCTGCTGAGGGA 350  
Db 425 CNTTYGTNCARCCNGTNTGYTTCNCCNAAAYATHAARTTYAARCCNWSNATHGWNATGT 484  
QY 351 TAGAAGTAATGGGTC 366  
Db 485 GYTGCGTATGCGTNG 500

## RESULT 6

US-09-134-000C-3076/c  
; Sequence 3076, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3076  
; LENGTH: 1242  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-3076

Query Match 3.6%; Score 34.8; DB 4; Length 1242;

Best Local Similarity 52.0%; Pred. No. 0.71; 72; Indels 0; Gaps 0;

Matches 78; Conservative 0; Mismatches 72

QY 675 AACACCAGTAATGCTTACATTTTGAAGTATGGAAGGCAACATTTGGTGGAGTATGTTGC 734

Db 156 AACATCAGACGATTCCTAATAGAGATCGTTAACAAAACGTTGTGAATATCTTGTAC 97

QY 735 ATTAGAATAAGACCCGATCAAGAGGGAACAGTGAATCCATCGGAGCATTTAT 794

Db 96 ATTGCCATAATATCTTCAATCAGCCAGCTTTTCCAAAAATCCATAGACTGCATTGA 37

QY 795 AAGTGTTCAGCCCAACATTTCAAGTTTACC 824

Db 36 AAGTCTATCTCTTAGCTTCAAAATCTCACC 7

## RESULT 7

US-09-620-312D-277/c

; Sequence 277, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunging

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. 6569662el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 784CIP2B

; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt\_FL\_genes Version 1.0

; SEQ ID NO 277

; LENGTH: 3146

; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (23)...(3130)  
US-09-620-312D-277

Query Match 3.6%; Score 34.8; DB 4; Length 3146;  
Best Local Similarity 46.1%; Pred. No. 1.2;  
Matches 117; Conservative 0; Mismatches 137; Indels 0; Gaps 0;  
QY 287 CGTACAATTAAACGAGCGGCTCTTCACTAGTCCAACTTGAACACACTTGAACACACTTCTTGA 346  
DB 1503 CTTAAGAAAACGAGCGAGCATGTCTAGAGTCTGGAACCTGCGAAGTTTATTTTCA 1444  
QY 347 GGGATAGAAATAATTGGGTGCATGATACCCCGAAATATTTCTATGGAAACAAGCCATGGA 406  
DB 1443 AGATTGATTAATCAATGGGTGCAGAGAGCTGGGCATCTCAGTCACACACCGGTCTTGCT 1384  
QY 407 ATGCAAACTAGCAACTGATGGCCCAATPACCAATTACCCAGTAAAGTTTCAAACTTAACAG 466  
DB 1383 CTGTAACAGGAGACTTTTCCTGTTCCTCAATGCACCAACAATCAAACTCTGGGTAAATCATTC 1324  
QY 467 ACTTCTACTTAACATCTCCTATATAAATTTGAGCCCAAGACGGCTGCCAATTAACCTCG 526  
DB 1323 TTGCTATGTCTGAATCTCTCTTAAAGTAAAGCAATCCGGAAGGACAGCATGTGAG 1264  
QY 527 CAATAGAACTCTGG 540  
DB 1263 GCCACACTCCTTG 1250

RESULT 8  
US-09-601-198-60/c  
Sequence 60, Application US/09601198  
Patent No. 65311583  
GENERAL INFORMATION:  
APPLICANT: Casseil, Gail H.  
APPLICANT: Chen, Ellison Y.  
APPLICANT: Glass, Jennifer S.  
APPLICANT: Glass, John I.  
APPLICANT: Heiner, Cheryl R.  
APPLICANT: Lefkowitz, Elliot  
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA  
FILE REFERENCE: UAB-13452/22  
CURRENT APPLICATION NUMBER: US/09/601,198  
CURRENT FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/073,189  
PRIOR FILING DATE: 1998-01-30  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 60  
LENGTH: 15016  
TYPE: DNA  
ORGANISM: Ureaplasma urealyticum  
US-09-601-198-60

Query Match 3.6%; Score 34.8; DB 4; Length 15016;  
Best Local Similarity 50.6%; Pred. No. 2.6;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 652 GTAGTCCCAATAAATAGTTAAGGACACACAGTAATGTACATTTGAATGGAAGCA 711  
DB 5691 GTTGGGCTGTATCAGTTAATAAACCAGCAAGGAATTTGATATTCACAGCA 5632  
QY 712 AACATTGGTGGAGTAGTTGCAATTAGATAAAGCCCCCAATCAAGAGGGACAGTG 771  
DB 5631 AATCAGGTTATGATAATTTTATTATTAATAAGATATTAATAATCAAAATATCTTGTA 5572  
QY 772 ACAATTCCATACGGAGCATTTATAGTGTTCAGCAACATTTCAA 817  
DB 5571 GGAAACATATTTAAAGTTAAAGTATTAATGTTGCACACCCCTTCAA 5526

RESULT 9  
US-09-328-352-1845  
Sequence 1845, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Galy L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: CTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 1845  
LENGTH: 1161  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-1845

Query Match 3.6%; Score 34.2; DB 4; Length 1161;  
Best Local Similarity 54.3%; Pred. No. 1.1;  
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
QY 717 TGGTTGGAGTAGTTGCAATTTAGATAAGACCCCAATCAAGAGGGACAGTGACAT 776  
DB 78 TGGTGTGGATATTGTTGGCGCTTCATTGAGCCCTTACAGCTTCATGCCACACTTACAGT 137  
QY 777 TCCATACGGAGCATTTTATAAGTGTTCAGGCCAACATTTCAAGCTTACCAAAATTACACAGA 836  
DB 138 TCCATTGGAACCGAGTTACGTGATGAGCTTATGGCATTAACCTTTACCCGATGACAATGA 197  
QY 837 ACTTTAC 843  
DB 198 AATTGAC 204

RESULT 10  
US-08-916-421B-1  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bult et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco  
Patent No. 6503729  
TITLE OF INVENTION: jannaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jannaschii  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature

```

; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)

```

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

```

Query Match 3.5%; Score 33.8; DB 4; Length 1664976;  
 Best Local Similarity 53.4%; Pred. No. 59;  
 Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 732 TGCATTAGTAATAAGACCCCAATCAAGAGGGAACAGTGCACAAATTCCTACACGAGCAAT 791  
Db 274434 TTACATATGGTATAGGTAAAGGAAATCAGGACCAATAATCTTTCACATCTCTCAT 274493  
QY 792 TATAAGTGTGACGCCCAACTTCAAGCTTACCAAAATACACAGAACTTTTACTTAGAGGA 851  
Db 274494 TATAATGTTTGAAGCTTTCATTTGTTGATAAACTATTAAATCAGTTTTTAAATTTATATCT 274553  
QY 852 CGTGGAGATTGA 864  
Db 274554 CGTGTAGTTTGA 274566

## RESULT 11

US-09-224-024-27/c  
; Sequence 27, Application US/09224024  
; Patent No. 6056353  
; GENERAL INFORMATION:  
; APPLICANT: Leslie Hickie  
; APPLICANT: Jewel Payne  
; TITLE OF INVENTION: Materials and Methods for the Control of  
; TITLE OF INVENTION: Calliphoridae Pests  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/224,024  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/856,226  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA79  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3543 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-224-024-27

Query Match 3.5%; Score 33.4; DB 3; Length 3543;  
Best Local Similarity 46.1%; Pred. No. 3.4;  
Matches 148; Conservative 0; Mismatches 171; Indels 2; Gaps 1;  
QY 439 TTACCAGTAAGTTTCAACCTACAGACTTCTATCAACAATCTCTTAACTTGA 498  
Db 1734 TTGACATGTAATTTTGAATGATCTTTGAATCAATTAATCCCTCTGTATGACCAGG 1675  
QY 499 CCCAAGAACGGCTGCCCAATTAACCTCGCAATAGAACTCTGTTAAGCAGAGAGCTTGG 558  
Db 1674 TCCTTGAACAACCTTAGAAGCA--GTCCCAAGTGAATTCGCTTTTACAGCTGGAATTTGG 1617  
QY 559 AGRACAACAGGAATTAACAGCGATGACGACAGAGTAATGATATGATGATTAATTAATGACGGA 618  
Db 1616 GTAGTAAATGTGTATAAATTTGTTATTTTAGATCAACACTAGAGTGTGTGCCAAGCAAC 1557

QY 619 TTACAACGGCTGGCTCCAAAGTTAAGAGAGATTGTAGTCCCAATTAATAGTTAAACGGAACA 678  
Db 1556 GTATACACTTGAAGTTTATATGTTGCAGGATACCTAAGACTTTTAAATAAATGATAAATA 1497  
QY 679 COAGTAAATGCTACATTTGAAGTATGGAAGGCAAAACATTTGGTGGAGATATGTTGCAATTT 738  
Db 1496 TGACTATAGTTATCATATGTTGAAAAAGGGTAGGGATTGCTTGATTCTCTCTCTGTTTA 1437  
QY 739 AGAATAAAGACCCCAATCAA 759  
Db 1436 AGAATTGTAACCCGAAATA 1416

## RESULT 12

PCT-US94-07902-27/c  
; Sequence 27, Application PC/TUS9407902  
; GENERAL INFORMATION:  
; APPLICANT: Street address: 4980 Carroll Canyon Road  
; APPLICANT: City: San Diego  
; APPLICANT: State/Province: California  
; APPLICANT: Country: US  
; APPLICANT: Postal code/Zip: 92121  
; APPLICANT: Phone number: (619) 453-8030 Fax number: (619) 453-5991  
; APPLICANT: Telex number:  
; TITLE OF INVENTION: Materials and Methods for the Control of  
; TITLE OF INVENTION: Calliphoridae Pests  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07902  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA79  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3543 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
PCT-US94-07902-27

Query Match 3.5%; Score 33.4; DB 5; Length 3543;  
Best Local Similarity 46.1%; Pred. No. 3.4;  
Matches 148; Conservative 0; Mismatches 171; Indels 2; Gaps 1;  
QY 439 TTACCAGTAAGTTTCAACCTACAGACTTCTATCAACAATCTCTTAACTTGA 498  
Db 1734 TTGACATGTAATTTTGAATGATCTTTGAATCAATTAATCCCTCTGTATGACCAGG 1675  
QY 499 CCCAAGAACGGCTGCCCAATTAACCTCGCAATAGAACTCTGTTAAGCAGAGAGCTTGG 558  
Db 1674 TCCTTGAACAACCTTAGAAGCA--GTCCCAAGTGAATTCGCTTTTACAGCTGGAATTTGG 1617

QY 559 AGAACACAGGAAATTAACAGCGATGAGCAAGAGTAATGATATGATGATTTACTATGACGGA 618  
Db 1616 GTAGTAAATGTGTATAAATTTGATTTTATTTAGCATCAACACTAGAGTGTGTCCAAAGCAAC 1557  
QY 619 TTACAAACGGCTGGCTCCAAAGTAAAGAGAGATTGTAGTCCCAATAATAGTTTAAACGGAACA 678  
Db 1556 GTATACACTTGAGTTTATATATGTTGCAGGATACTAGACTTTTAAATAATGATATAATA 1497  
QY 679 CAGTAAATGCTACATTTGAAGTATGGAAGCAACATGTTGGAGATATGTTGCAATTT 738  
Db 1496 TGACTATAGTTATCATATGTTGGAAGAGGCTAGGAGTTGCTGATTTCTCTCTGTTTA 1437  
QY 739 AGAATAAAGACCCCAATCAA 759  
Db 1436 AGAATTGTAACCCGAAATA 1416

## RESULT 13

US-09-385-982-400  
; Sequence 400, Application US/09385982  
; Patent No. 6262334  
; GENERAL INFORMATION:  
; APPLICANT: ENDEGE, WILSON O., ET AL.  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCNA-260XX  
; CURRENT APPLICATION NUMBER: US/09/385,982  
; CURRENT FILING DATE: 1999-08-30  
; EARLIER APPLICATION NUMBER: 09/328,111  
; EARLIER FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: 60/117,393  
; EARLIER FILING DATE: 1999-01-27  
; EARLIER APPLICATION NUMBER: 60/098,639  
; EARLIER FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 544  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 400  
; LENGTH: 612  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(612)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-385-982-400

Query Match 3.5%; Score 33.2; DB 3; Length 612;  
Best Local Similarity 52.1%; Pred. No. 1.5;  
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
QY 450 AGTTTCAAACTTAACAGACTTCTATCTTAACAATCTCTTAACTTGAAGCTTGAAGCAAGG 509  
Db 30 AGTTTCAAGCTTCAACATATTCATGTCATCAAGCTTTTCCAAAGATCAGGCTTAAAGAAA 89  
QY 510 CTGCGCAATTAACITCGCAATAGATCTGTTAAACGAGAGCTTTGGAGCAACAGG 569  
Db 90 CAGATCAAAAGCTTTTCATCATCTCTGTAATAACGAGAGAGTTTAAAGAACAGAGAA 149  
QY 570 AATTAACAGCGATGAGCAAGAA 591  
Db 150 AAGCAACAGTGAATCTGAAGAA 171

## RESULT 14

US-09-621-976-2813/c  
; Sequence 2813, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 2813  
; LENGTH: 832  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: CDS  
; LOCATION: 235..399  
US-09-621-976-2813

Query Match 3.4%; Score 32.8; DB 4; Length 832;  
Best Local Similarity 9.4%; Pred. No. 2.4;  
Matches 28; Conservative 139; Mismatches 131; Indels 0; Gaps 0;

QY 406 AATGCAAACTAGCCAACTGATGCCCAATACCAATACCCAGTAAAGTTTCAAACCTAACA 465  
Db 308 MKKEMSTEWYCYMCKMCCMYGRCAWYTWARGRWYSYANGKWKSMESAMCTRMYYK 249  
QY 466 GACTTCTATCAAACTCTCTATAACTTGAGCCCAAGACGGCTGCCAATTAACCTC 525  
Db 248 KGSYTWTKCTCATWYKWKYKRWKSKTCSGSRGGTMTSYSTRSYMYWASWYTM 189  
QY 526 GCAATAGAAATCTGGTTAACGAGAGAAGCTTGGAGAACACACAGGAATTAACAGCGATGAG 585  
Db 188 CWWGGRWSTYWMANGKWKWYATTTWRAMWWAAWTKMYWWKMSRGRGAAYR 129  
QY 586 CAAGAAGTAATGATATGATTTACTATGACGATTAACACCGGCTGGCTCCAAAGCTTAAG 645  
Db 128 TMMMGYRYWWRKSKYRRTRCAWAYAWTKRSYKWKWKRCMWMWMMAMAYGKTMM 69  
QY 645 GAGATTGTAGTCCCAATATAGTTAAACGGAACACAGTAAATGCTACATTTGAAGTAT 703  
Db 68 RACKTKRYWRWAWANWRWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTW 11

## RESULT 15

US-08-319-704-5  
; Sequence 5, Application US/08319704  
; Patent No. 5814617  
; GENERAL INFORMATION:  
; APPLICANT: Hofman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard C.  
; APPLICANT: Doonan, Denise L.  
; TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and  
; TITLE OF INVENTION: Erythrocytic Stage Immunogen and Gene  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Naval Medical R & D Command  
; STREET: Bldg 1, T-12, 8901 Wisconsin Avenue  
; CITY: Bethesda  
; STATE: Maryland  
; COUNTRY: U.S.A  
; ZIP: 20889-5606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/319,704  
; FILING DATE: 07-OCT-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. David Spevack  
; REGISTRATION NUMBER: 24,743  
; REFERENCE/DOCKET NUMBER: 75,206  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-1022



```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1422 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-319-704-5

```

```

Query Match      3.4%; Score 32.8; DB 1; Length 1422;
Best Local Similarity 50.0%; Pred. No. 3.2;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 675 AACCCAGTAAATGCTACATTGAAAGTATCGAAGGCAAAACATTGGTTGGAGATATGTTGC 734
Db 448 AAAACAAGTATGGTAAATAATGGCAAAATATGGCTCCCAAAATGTTATTAGAGTAAGTTGA 507

Qy 735 ATTTAGATTAAGACCCCAATCAAGAGAGGGAACAGTGCACATTCCTACCGAGCATTAT 794
Db 508 TTAAAGTAAATATATATCTAATTTAAGTATCCATATATACATGTTTATACCTTTTGAGTTAT 567

Qy 795 AAGTGTTCGACCCCAACATTTCAAGCTTACCAAAATTACACAGAAC 838
Db 568 ACGTAAATGTATATACGAACTTGTATTTAATATTTTAAAAAAC 611

```

Search completed: July 2, 2004, 00:48:11  
Job time : 107 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 20:13:10 ; Search time 52 Seconds  
(without alignments)  
1734.288 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701  
Sequence: 1 MSKKFVIVSILATLLVQAI.....HLEWITNITLTPLDRPLIS 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	13	US-10-003-759-5
2	1701	100.0	319	13	US-10-121-032-64
3	1701	100.0	319	13	US-10-093-037-64
4	1617	95.1	300	14	US-10-228-063-55
5	464	27.3	274	15	US-10-369-493-3041
6	166.5	9.8	261	13	US-10-003-759-2
7	160.5	9.4	225	14	US-10-294-444-1
8	128	7.5	260	12	US-10-441-625-23
9	128	7.5	260	14	US-10-441-626-23
10	120	7.1	1228	10	US-09-917-383-1
11	120	7.1	1228	10	US-09-917-383-1
12	119.5	7.0	553	9	US-09-888-224-2
13	116	6.8	1043	10	US-09-917-384-6
14	116	6.8	1043	10	US-09-917-383-6
15	114.5	6.7	381	12	US-10-441-625-22

16	114.5	6.7	381	14	US-10-441-626-22	Sequence 22, Appl
17	112.5	6.6	264	12	US-10-441-625-24	Sequence 24, Appl
18	112.5	6.6	264	14	US-10-441-626-24	Sequence 24, Appl
19	111.5	6.6	269	14	US-10-156-761-10105	Sequence 10105, A
20	108.5	6.4	221	10	US-09-917-384-13	Sequence 13, Appl
21	108.5	6.4	221	10	US-09-917-383-13	Sequence 13, Appl
22	108	6.3	371	9	US-09-739-861A-1	Sequence 1, Appl
23	108	6.3	371	9	US-09-795-583-1	Sequence 1, Appl
24	108	6.3	371	12	US-10-441-625-21	Sequence 21, Appl
25	108	6.3	371	14	US-10-441-626-21	Sequence 21, Appl
26	108	6.3	386	9	US-09-739-861A-5	Sequence 5, Appl
27	108	6.3	386	9	US-09-795-583-5	Sequence 5, Appl
28	107.5	6.3	231	10	US-09-917-384-7	Sequence 7, Appl
29	107.5	6.3	231	10	US-09-917-383-7	Sequence 7, Appl
30	107	6.3	228	10	US-09-917-383-14	Sequence 14, Appl
31	107	6.3	228	10	US-09-917-383-14	Sequence 14, Appl
32	103	6.1	708	12	US-09-848-909-29	Sequence 29, Appl
33	103	6.1	842	14	US-10-132-350-50	Sequence 50, Appl
34	103	6.1	842	14	US-10-132-350-52	Sequence 52, Appl
35	103	6.1	1338	12	US-09-848-909-35	Sequence 35, Appl
36	102.5	6.0	533	15	US-10-369-493-6113	Sequence 6113, Ap
37	102.5	6.0	533	15	US-10-369-493-6114	Sequence 6114, Ap
38	100.5	5.9	345	9	US-09-775-195-4	Sequence 4, Appl
39	100.5	5.9	345	14	US-10-155-947-4	Sequence 4, Appl
40	100.5	5.9	345	14	US-10-094-407A-4	Sequence 4, Appl
41	100	5.9	881	9	US-09-850-351A-32	Sequence 32, Appl
42	100	5.9	881	15	US-10-452-002A-8	Sequence 8, Appl
43	99	5.8	259	12	US-10-441-625-5	Sequence 5, Appl
44	99	5.8	259	14	US-10-441-626-5	Sequence 5, Appl
45	99	5.8	4620	12	US-10-282-122A-68921	Sequence 68921, A

#### ALIGNMENTS

#### RESULT 1

US-10-003-759-5  
; Sequence 5, Application US/10003759  
; Publication No. US20020102699A1  
; GENERAL INFORMATION:  
; APPLICANT: Wicher, Kryzysstof B.  
; APPLICANT: Holst, Olof Peder  
; APPLICANT: Hachem, Maher Youssef Abou  
; APPLICANT: Karlsson, Eva Margareta No. US20020102699A1dberg  
; APPLICANT: Hreggvidsson, Gudmundur O.  
; TITLE OF INVENTION: Thermostable Cellulase  
; FILE REFERENCE: P5099PC00  
; CURRENT APPLICATION NUMBER: US/10/003,759  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: PCT/IS01/00012  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 09/594,884  
; PRIOR FILING DATE: 2000-06-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(19)  
; OTHER INFORMATION:  
; NAME/KEY: DOMAIN  
; LOCATION: (28)..(49)  
; OTHER INFORMATION: Linker Moiety  
; NAME/KEY: DOMAIN  
; LOCATION: (50)..(319)  
; OTHER INFORMATION: Catalytic Domain  
US-10-003-759-5

Query Match 100.0%; Score 1701; DB 13; Length 319;  
Best Local Similarity 100.0%; Pred. No. 7.9e-160;

Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSSTPPQTLSTTKVLKIRYPDDG 60  
 Db 1 MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSSTPPQTLSTTKVLKIRYPDDG 60

QY 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120  
 Db 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

QY 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPIFAIESW 180  
 Db 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPIFAIESW 180

QY 181 LTREAWRTTGINSDEQVMIWYYDGLQPGAGSKVEIWPVIIVNGTVPNATFEVWKANIG 240  
 Db 181 LTREAWRTTGINSDEQVMIWYYDGLQPGAGSKVEIWPVIIVNGTVPNATFEVWKANIG 240

QY 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTLYLEDVEIGTEGTPSTTSAH 300  
 Db 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTLYLEDVEIGTEGTPSTTSAH 300

QY 301 LEWMITNITLPLDRPLIS 319  
 Db 301 LEWMITNITLPLDRPLIS 319

RESULT 2  
 US-10-121-032-64  
 ; Sequence 64, Application US/10121032  
 ; Publication No- US20020155550A1

GENERAL INFORMATION:  
 ; APPLICANT: Bylina, Edward J.  
 ; TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
 ; NUMBER OF SEQUENCES: 72  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Gray Cary Ware & Freidenrich LLP  
 ; STREET: 4365 Executive Drive, Suite 1600  
 ; CITY: San Diego  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92121  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: Windows95  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/121,032  
 ; FILING DATE: 09-Apr-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/134,078  
 ; FILING DATE: 13-AUG-1998  
 ; APPLICATION NUMBER: 08/949,026  
 ; FILING DATE: 10-OCT-1997  
 ; APPLICATION NUMBER: 60/056,916  
 ; FILING DATE: 06-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 09010/024002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 858/677-1456  
 ; TELEFAX: 858/677-1465  
 ; INFORMATION FOR SEQ ID NO: 64:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 319 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: internal  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 64:

US-10-121-032-64

Query Match 100.0%; Score 1701; DB 13; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-160;  
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSSTPPQTLSTTKVLKIRYPDDG 60  
 Db 1 MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSSTPPQTLSTTKVLKIRYPDDG 60

QY 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120  
 Db 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

QY 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPIFAIESW 180  
 Db 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPIFAIESW 180

QY 181 LTREAWRTTGINSDEQVMIWYYDGLQPGAGSKVEIWPVIIVNGTVPNATFEVWKANIG 240  
 Db 181 LTREAWRTTGINSDEQVMIWYYDGLQPGAGSKVEIWPVIIVNGTVPNATFEVWKANIG 240

QY 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTLYLEDVEIGTEGTPSTTSAH 300  
 Db 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTLYLEDVEIGTEGTPSTTSAH 300

QY 301 LEWMITNITLPLDRPLIS 319  
 Db 301 LEWMITNITLPLDRPLIS 319

RESULT 3  
 US-10-093-037-64  
 ; Sequence 64, Application US/10093037  
 ; Publication No- US20030078397A1

GENERAL INFORMATION:  
 ; APPLICANT: Jay M. Short  
 ; APPLICANT: Bylina, Edward  
 ; APPLICANT: Swanson, Ronald V.  
 ; APPLICANT: Mathur, Eric J.  
 ; APPLICANT: Lam, David E.  
 ; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
 ; FILE REFERENCE: 09010-024006  
 ; CURRENT APPLICATION NUMBER: US/10/093,037  
 ; CURRENT FILING DATE: 2002-03-06  
 ; PRIOR APPLICATION NUMBER: US 09/910,579  
 ; PRIOR FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: US 09/134,078  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 08/949,026  
 ; PRIOR FILING DATE: 1997-10-10  
 ; PRIOR APPLICATION NUMBER: US 60/056,916  
 ; PRIOR FILING DATE: 1996-12-06  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 64  
 ; LENGTH: 319  
 ; TYPE: PRT  
 ; ORGANISM: Pyrococcus furiosus  
 ; US-10-093-037-64

Query Match 100.0%; Score 1701; DB 14; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-160;  
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSSTPPQTLSTTKVLKIRYPDDG 60  
 Db 1 MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSSTPPQTLSTTKVLKIRYPDDG 60

QY 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120  
 Db 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

QY 121 WHGYPETFYGNKPNWANYATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIESW 180  
DB 121 WHGYPETFYGNKPNWANYATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIESW 180  
QY 181 LTRAWRTTGINSDQEWIMWYVDGLQPGSKVKEIWPVIIVNGTTPVNAFEEVKANIG 240  
DB 181 LTRAWRTTGINSDQEWIMWYVDGLQPGSKVKEIWPVIIVNGTTPVNAFEEVKANIG 240  
QY 241 WEYVAFRIKPIKEGTVTIPYGFISVAANISSLPNTYLYLEDEVEIGTFGTPSTTSAH 300  
DB 241 WEYVAFRIKPIKEGTVTIPYGFISVAANISSLPNTYLYLEDEVEIGTFGTPSTTSAH 300  
QY 301 LEWMITNITLPLDRPLIS 319  
DB 301 LEWMITNITLPLDRPLIS 319

## RESULT 4

US-10-228-063-55

; Sequence 55, Application US/10228063

; Publication No. US20030135868A1

; GENERAL INFORMATION:

; APPLICANT: Lanahan, Mike

; TITLE OF INVENTION: Self-processing Plants and Plant Parts

; FILE REFERENCE: 109846.317

; CURRENT APPLICATION NUMBER: US/10/228.063

; CURRENT FILING DATE: 2002-12-12

; NUMBER OF SEQ ID NOS: 60

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 55

; LENGTH: 300

; TYPE: PRT

; ORGANISM: Pyrococcus furiosus

US-10-228-063-55

Query Match 95.1%; Score 1617; DB 14; Length 300;

Best Local Similarity 100.0%; Pred. No. 1.5e-151;

Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 IYFVEKHTSDKSTNTSSPPQTTSTTKVLKIRYPDDGWPAGPIDKXGDNPEFYI 79  
DB 1 IYFVEKHTSDKSTNTSSPPQTTSTTKVLKIRYPDDGWPAGPIDKXGDNPEFYI 60

QY 80 EINLWNILNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPYFYGKPNWANY 139  
DB 61 EINLWNILNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPYFYGKPNWANY 120

QY 140 ATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIESWLTREAWRTTGINSDQEWIM 199  
DB 121 ATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIESWLTREAWRTTGINSDQEWIM 180

QY 200 IWYVDGLQPGSKVKEIWPVIIVNGTTPVNAFEEVKANIGWEYVAFRIKPIKEGTVTI 259  
DB 181 IWYVDGLQPGSKVKEIWPVIIVNGTTPVNAFEEVKANIGWEYVAFRIKPIKEGTVTI 240

QY 260 PYGAFISVAANISSLPNTYLYLEDEVEIGTFGTPSTTSAHLEWMITNITLPLDRPLIS 319  
DB 241 PYGAFISVAANISSLPNTYLYLEDEVEIGTFGTPSTTSAHLEWMITNITLPLDRPLIS 300

## RESULT 5

US-10-369-493-3041

; Sequence 3041, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 3041

; LENGTH: 274

; TYPE: PRT

; ORGANISM: Thermotoga maritima

US-10-369-493-3041

Query Match 27.3%; Score 464; DB 15; Length 274;

Best Local Similarity 40.4%; Pred. No. 2.6e-37;

Matches 93; Conservative 43; Mismatches 84; Indels 10; Gaps 4;

QY 79 IEINLWNILNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPYFYGKPNWANY 138  
DB 39 MELNFWNKSYEG--ETWLKFDGKVEFYADLYNVLQNPDSWVHGYPYFYGKPNWAGH 96

QY 139 YATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIESWLTREAWRTTGINSDQEWIM 198  
DB 97 --NSGVEFLPVKVKDLPDFYVTLDSYTWENNLPINLAWETWITRSPDOTS--VSSGDABI 153

QY 199 MIWYVDGLQPGSKVKEIWPVIIVNGTTPVNAFEEVKANIGWEYVAFRIKPIKEGTVTI 258  
DB 154 MWYFNNVLMPGGQKVDFTTVEINGVKQETKWDVYPAPWGDYLAPELTTPEMKEGVKV 213

QY 259 IYPGAFISVAANI-----SSLPNTYLYLEDEVEIGTFGTPSTTSAHLEW 303  
DB 214 INVKDFVQKAAEVVKHSTRIDNFEELYFCVWEIGTEFGDPTNTAAKFGW 263

## RESULT 6

US-10-003-759-2

; Sequence 2, Application US/10003759

; Publication No. US20020102699A1

; GENERAL INFORMATION:

; APPLICANT: Holst, Olof Peder

; APPLICANT: Wicher, Krysztof B.

; APPLICANT: Hachem, Maher Youssef Abou

; APPLICANT: Karlsson, Eva Margareta No. US20020102699A1dberg

; APPLICANT: Hreggsvidsson, Gudmundur O.

; TITLE OF INVENTION: Thermostable Cellulase

; FILE REFERENCE: P5099PC00

; CURRENT APPLICATION NUMBER: US/10/003,759

; CURRENT FILING DATE: 2001-10-23

; PRIOR APPLICATION NUMBER: PCT/IS01/00012

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: 09/594,884

; PRIOR FILING DATE: 2000-06-15

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 261

; TYPE: PRT

; ORGANISM: Rhodothermus marinus

US-10-003-759-2

Query Match 9.8%; Score 166.5; DB 13; Length 261;

Best Local Similarity 26.6%; Pred. No. 7.3e-08;

Matches 66; Conservative 35; Mismatches 108; Indels 39; Gaps 11;

QY 59 DGEWPGAPIDKXGDNPEFYIEI-----NLWNILNATGFAEMTYNLTSGVLHYV 107  
DB 19 DMLFPDGDNGKPEPEPEPTVELCGRDWARDVAGGRYRVINNWGAQAQCEVGL---- 74

QY 108 QQLDNIVL---RDRSNWVHGYPYFYGKPNWANYATDGPILPSPKSVNLTDFYLTISY 163  
DB 75 -ETGNFTITRADHNGNVAAPALYFGCE-WGACTSNG---LPRVOELSD--VRTSW 127

QY 164 KLEPKNGLPINFAIESWLTREAWRTTGINSDQEWIMWYVDGLQPGSKVKEIWPVIIV 222  
DB 126 TLTPTTGRWNAAYDIWFSPVTNSNGY--SGGAELMILNNGVMPGSGSRVATVEL--- 183

QY 223 VNGTVPNATFEVWKNIGWEYVAFRIKTIPIKEGTVTIPYGFAPISVAANISLPLNYTELYL 282  
 Db 184 -----AGATWEVWYADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--RPEWYL 235  
 QY 283 EDVEIGTE 290  
 Db 236 HAVETGFE 243

RESULT 7  
 US-10-294-444-1  
 ; Sequence 1, Application US/1029444  
 ; Publication No. US20030199072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Crennell, Susan J.  
 ; APPLICANT: Karlsson, Eva M. N.  
 ; APPLICANT: Hreggvidsson, Gudmundur O.  
 ; APPLICANT: Kristjansson, Jakob K.  
 ; APPLICANT: Aevansson, Arnthor  
 ; TITLE OF INVENTION: CRYSTAL AND STRUCTURE OF A THERMOSTABLE  
 ; TITLE OF INVENTION: GLYCOSOL HYDROLASE AND USE THEREOF, AND MODIFIED PROTEINS  
 ; FILE REFERENCE: 2739.2009-000  
 ; CURRENT APPLICATION NUMBER: US/10/294,444  
 ; CURRENT FILING DATE: 2002-11-14  
 ; PRIOR APPLICATION NUMBER: IS 6353  
 ; PRIOR FILING DATE: 2002-04-19  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 225  
 ; TYPE: PRT  
 ; ORGANISM: Rhodothermus marinus  
 ; US-10-294-444-1

Query Match 9.4%; Score 160.5; DB 14; Length 225;  
 Best Local Similarity 31.4%; Pred. No. 2.3e-07;  
 Matches 55; Conservative 23; Mismatches 78; Indels 19; Gaps 8;  
 QY 117 DRSNVHGYPEIFYGNKPNWANYATDGPILPSKYSNLTDFLTISYKLEPKNGLPINFA 176  
 Db 51 DNGNNVAAPAIYFGCH-WGACTNSG--LPRRVOELSD--VRTSWTLTBITGRWNA 104  
 QY 177 ISWLTREAWRTTGINSDEQVMIWYD-GLQAGSKVKSLVVPILVNGTVPVNATFEVW 235  
 Db 105 YDIWSPVTSNGY-SGAELMIWLNNWGVMPGGSRVATVEL-----AGATWEVW 155  
 QY 236 KANIGWEYVAFRIKTIPIKEGTVTIPYGFAPISVAANISLPLNYTELYLEDEVEIGTE 290  
 Db 156 YADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--RPEWYLHAVETGFE 207

RESULT 8  
 US-10-441-625-23  
 ; Sequence 23, Application US/10441625  
 ; Publication No. US20030203467A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gualfetti, Peter  
 ; APPLICANT: Mitchinson, Colin  
 ; APPLICANT: Phillips, Jay Ian  
 ; TITLE OF INVENTION: Novel Variant EGIII-Like Cellulase  
 ; TITLE OF INVENTION: Compositions  
 ; FILE REFERENCE: GC631  
 ; CURRENT APPLICATION NUMBER: US/10/441,625  
 ; CURRENT FILING DATE: 2003-05-19  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 23  
 ; LENGTH: 260  
 ; TYPE: PRT  
 ; ORGANISM: Rhodothermus marinus  
 ; US-10-441-625-23

Query Match 7.5%; Score 128; DB 12; Length 260;  
 Best Local Similarity 24.1%; Pred. No. 0.00048;  
 Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;  
 QY 59 DGEWPGAPDKDGDGNPEFYIEI-----NLWNILNATGPAEMTYNLTSGVLHYV 107  
 Db 19 DWLFDPDNGKPEPEPTVELCGRWDAVDAGGRYRVINNVWGAETAQCIIEVGL---- 74  
 QY 108 QOLDNIVL-----RDRSNVHGYPEIFYGNKPNWANYATDGPILPSKYSNLTDFLTISY 163  
 Db 75 -ETGNFTITRADHDNGNVAAPAIYFGCH-WAPARAIRDCAARAGAVRAHELDVT--- 129  
 QY 164 KLEPKNGLP-----NFAIESWLTREAWRTTGINSDEQVMIWYD-GLQAGSKVKEI 217  
 Db 130 -----PITGRWNAAYDIWFSPTNSNGY-SGAELMIWLNNWGVMPGGSRVATV 180  
 QY 218 VVPIIVNGTVPVNATFEVWKNIGWEYVAFRIKTIPIKEGTVTIPYGFAPISVAANISLPLNY 277  
 Db 181 EL-----AGATWEVWYADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--R 229  
 QY 278 TELYLEDEVEIGTE 290  
 Db 230 PEWYLHAVETGFE 242

RESULT 9  
 US-10-441-626-23  
 ; Sequence 23, Application US/10441626  
 ; Publication No. US20030186418A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gualfetti, Peter  
 ; APPLICANT: Mitchinson, Colin  
 ; APPLICANT: Phillips, Jay Ian  
 ; TITLE OF INVENTION: No. US20030186418A1el Variant EGIII-Like Cellulase  
 ; TITLE OF INVENTION: Compositions  
 ; FILE REFERENCE: GC631  
 ; CURRENT APPLICATION NUMBER: US/10/441,626  
 ; CURRENT FILING DATE: 2003-05-19  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 23  
 ; LENGTH: 260  
 ; TYPE: PRT  
 ; ORGANISM: Rhodothermus marinus  
 ; US-10-441-626-23

Query Match 7.5%; Score 128; DB 14; Length 260;  
 Best Local Similarity 24.1%; Pred. No. 0.00048;  
 Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;  
 QY 59 DGEWPGAPDKDGDGNPEFYIEI-----NLWNILNATGPAEMTYNLTSGVLHYV 107  
 Db 19 DWLFDPDNGKPEPEPTVELCGRWDAVDAGGRYRVINNVWGAETAQCIIEVGL---- 74  
 QY 108 QOLDNIVL-----RDRSNVHGYPEIFYGNKPNWANYATDGPILPSKYSNLTDFLTISY 163  
 Db 75 -ETGNFTITRADHDNGNVAAPAIYFGCH-WAPARAIRDCAARAGAVRAHELDVT--- 129  
 QY 164 KLEPKNGLP-----NFAIESWLTREAWRTTGINSDEQVMIWYD-GLQAGSKVKEI 217  
 Db 130 -----PITGRWNAAYDIWFSPTNSNGY-SGAELMIWLNNWGVMPGGSRVATV 180  
 QY 218 VVPIIVNGTVPVNATFEVWKNIGWEYVAFRIKTIPIKEGTVTIPYGFAPISVAANISLPLNY 277  
 Db 181 EL-----AGATWEVWYADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--R 229  
 QY 278 TELYLEDEVEIGTE 290  
 Db 230 PEWYLHAVETGFE 242

RESULT 10  
 US-09-917-384-1

; Sequence 1, Application US/09917384  
; Publication No. US2003009342A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS

; TITLE OF INVENTION: CELLULOGLYTICUS

; FILE REFERENCE: 40170.6US01

; CURRENT APPLICATION NUMBER: US/09/917,384

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1228

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of

; OTHER INFORMATION: Guxa

US-09-917-384-1

Query Match 7.1%; Score 120; DB 10; Length 1228;  
Best Local Similarity 22.7%; Pred. No. 0.028;  
Matches 75; Conservative 33; Mismatches 129; Indels 94; Gaps 16;

Qy 26 YHTSEDKSTNTSTPTQTLLTKIRYDDEGWPAGIDKDGNGP-----75

Db 822 YTVAAVDAAGNTSAPSTPVTAITTS-----PSPSTPTGTVTDTCPGNQNGVTSVQG 875

Qy 76 -EFYIEINLWN-----ILNATG-FAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHG 125

Db 876 DEYRVOTNEWSSAQOCLTINTATGAVTVSTANFSGGT-----GGAPATY 920

Qy 126 PEIFYGNKPWNAYATDGPILPSKVSNLTDFTYISYKLEPKNGLPINFAIESWLTREA 185

Db 921 PSYKKGCHWN---CTTKNVGMPIQISQ-----IGSAVTSWSTTQV 958

Qy 186 -----W-----RTTGINSDEQEVMIWI--YYDGLQPAGSKVKEIIVVPIVNGTPV 228

Db 959 SSGAYDVAYDIWNTSTPTTG-QPNGTETIMLNRSRGVQPFQSGTATGV-----TVA 1010

Qy 229 NATFEVWKA-NIGWEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNYTELYLEDVEI 287

Db 1011 GHTVNWQGGQTSWKIISY-VLTFGATSIISNLDLKAIFADAAARGSL--NTSDYLLDVEA 1067

Qy 288 GTEP--GTPSTTSAHLEWNIITLTPLDRP 316

Db 1068 GFETWQGGQGLGNSFVSVTSCTSPTSP 1098

RESULT 11

US-09-917-383-1

; Sequence 1, Application US/09917383

; Publication No. US2003010452A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS

; TITLE OF INVENTION: CELLULOGLYTICUS

; FILE REFERENCE: 40170.6US01

; CURRENT APPLICATION NUMBER: US/09/917,383

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1228

; TYPE: PRT

; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-1

Query Match 7.1%; Score 120; DB 10; Length 1228;

Best Local Similarity 22.7%; Pred. No. 0.028;

Matches 75; Conservative 33; Mismatches 129; Indels 94; Gaps 16;

Qy 26 YHTSEDKSTNTSTPTQTLLTKIRYDDEGWPAGIDKDGNGP-----75

Db 822 YTVAAVDAAGNTSAPSTPVTAITTS-----PSPSTPTGTVTDTCPGNQNGVTSVQG 875

Qy 76 -EFYIEINLWN-----ILNATG-FAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHG 125

Db 876 DEYRVOTNEWSSAQOCLTINTATGAVTVSTANFSGGT-----GGAPATY 920

Qy 126 PEIFYGNKPWNAYATDGPILPSKVSNLTDFTYISYKLEPKNGLPINFAIESWLTREA 185

Db 921 PSYKKGCHWN---CTTKNVGMPIQISQ-----IGSAVTSWSTTQV 958

Qy 186 -----W-----RTTGINSDEQEVMIWI--YYDGLQPAGSKVKEIIVVPIVNGTPV 228

Db 959 SSGAYDVAYDIWNTSTPTTG-QPNGTETIMLNRSRGVQPFQSGTATGV-----TVA 1010

Qy 229 NATFEVWKA-NIGWEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNYTELYLEDVEI 287

Db 1011 GHTVNWQGGQTSWKIISY-VLTFGATSIISNLDLKAIFADAAARGSL--NTSDYLLDVEA 1067

Qy 288 GTEP--GTPSTTSAHLEWNIITLTPLDRP 316

Db 1068 GFETWQGGQGLGNSFVSVTSCTSPTSP 1098

RESULT 12

US-09-888-224-2

; Sequence 2, Application US/09888224

; Patent No. US20020120118A1

; GENERAL INFORMATION:

; APPLICANT: Lam, D. et al.

; TITLE OF INVENTION: Endoglucanases

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

; ADDRESSEE: CECCHI, STEWART & OLSTEIN

; STREET: 6 BECKER FARM ROAD

; CITY: ROSELAND

; STATE: NEW JERSEY

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH DISKETTE

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: WORD PERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/888,224

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/430,669

; FILING DATE: 28-Oct-1999

; APPLICATION NUMBER: US/08/651,572

; FILING DATE: 22-MAY-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Charles J. Herron

; REGISTRATION NUMBER: 28,019

; REFERENCE/DOCKET NUMBER: 331400-48

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700

; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:



LENGTH: 553 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS: Unknown  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-09-888-224-2

Query Match 7.0%; Score 119.5; DB 9; Length 553;  
Best Local Similarity 25.1%; Pred. No. 0.0099;  
Matches 45; Conservative 22; Mismatches 51; Indels 61; Gaps 9;  
QY 33 STSNTSTPPQTLSTT-----KVLKIRYPDDGE 61  
DB 374 TTTTSTSTPPTTTTTTTPTTTTQTPPTTTTTPSNVPEIWNV-LPTSSQ 432  
QY 62 WPGAPDKDGN-----PFYIEINLWNL-----NAT-----GFAEMT 96  
DB 433 YEGTSVEVVCDTQCASSWVGAPNLWGVKIGNATMDPNVWGEDVYKTAQDGTGSK 492  
QY 97 YNLTSGLVHYVQOLNIVLRDRSNWVHGYPFIYGNKPNWANYATDGP-IPLSKVN 154  
DB 493 MEIRGVLLK-VTNLWNLNHPKYNIM-AYPEVIYGAKEW-GNQPINAPNFVLPKVSQ 548

RESULT 13  
US-09-917-384-6

Sequence 6, Application US/09917384  
Publication No. US20030096342A1  
GENERAL INFORMATION:  
APPLICANT: DING, SHI-YOU  
APPLICANT: ADNEY, WILLIAM S.  
APPLICANT: VINZANT, TODD B.  
APPLICANT: DECKER, STEPHEN R.  
APPLICANT: HIMMEL, MICHAEL E.  
TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
FILE REFERENCE: 40170.6US01  
CURRENT APPLICATION NUMBER: US/09/917,384  
CURRENT FILING DATE: 2001-07-28  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 1043  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Guxa  
US-09-917-384-6

Query Match 6.8%; Score 116; DB 10; Length 1043;  
Best Local Similarity 23.2%; Pred. No. 0.055;  
Matches 69; Conservative 31; Mismatches 105; Indels 92; Gaps 15;  
QY 34 TSNTSTPPQTLSTTKVLKIRYPDDGWPAGAPID-----KGDGNPEFYIE 80  
DB 678 TSFTDTGLAAGTAYTYTVAADAAGNTSAPSTPVDCTPGPNQGVTSVQDG-----EYRVQ 733  
QY 81 INLWN-----ILNATG-FAEMTYNLTSGVLHYVQOLNIVLRDRSNWVHGYPFIY 131  
DB 734 TNEWSSAQOCLTINTATGAWTVSTANFSGGT-----GGAPATYPSIYKG 778  
QY 132 NKPWNANYATDGPILPLSKVSNLTDPLYLTISYKLEPKNGPLINFAIESWLTREA----- 185  
DB 779 CHWGN-----CTTKNVGMPIQISQ-----IGSAVTSWSTQVSSGAYD 816  
QY 186 -----W-----RTTGINSDEQEVMTWI-YYDGLQAGSKVKEIVVPIIIVNGTPVNATREV 234  
DB 817 VAYDIWTNSTPTTTG-OPNGTEIMLWNSRGVQVPGSQATGV-----TVAGHTWNV 868  
QY 235 WKA-NIGWEYVAFRIKTPKEGTVTIPYGAFISVAANISLNPNTYLYEDVEIGTE 290  
DB 869 WQOQQTWNKIISY-VLTGATISINLDKAIFADAAARGSL--NTSDYLLDVEAGFE 922

RESULT 14

US-09-917-383-6  
Sequence 6, Application US/09917383  
Publication No. US20030104522A1  
GENERAL INFORMATION:  
APPLICANT: DING, SHI-YOU  
APPLICANT: ADNEY, WILLIAM S.  
APPLICANT: VINZANT, TODD B.  
APPLICANT: DECKER, STEPHEN R.  
APPLICANT: HIMMEL, MICHAEL E.  
TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
FILE REFERENCE: 40170.6US01  
CURRENT APPLICATION NUMBER: US/09/917,383  
CURRENT FILING DATE: 2001-07-28  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 1043  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Guxa  
US-09-917-383-6

Query Match 6.8%; Score 116; DB 10; Length 1043;  
Best Local Similarity 23.2%; Pred. No. 0.055;  
Matches 69; Conservative 31; Mismatches 105; Indels 92; Gaps 15;  
QY 34 TSNTSTPPQTLSTTKVLKIRYPDDGWPAGAPID-----KGDGNPEFYIE 80  
DB 678 TSFTDTGLAAGTAYTYTVAADAAGNTSAPSTPVDCTPGPNQGVTSVQDG-----EYRVQ 733  
QY 81 INLWN-----ILNATG-FAEMTYNLTSGVLHYVQOLNIVLRDRSNWVHGYPFIY 131  
DB 734 TNEWSSAQOCLTINTATGAWTVSTANFSGGT-----GGAPATYPSIYKG 778  
QY 132 NKPWNANYATDGPILPLSKVSNLTDPLYLTISYKLEPKNGPLINFAIESWLTREA----- 185  
DB 779 CHWGN-----CTTKNVGMPIQISQ-----IGSAVTSWSTQVSSGAYD 816  
QY 186 -----W-----RTTGINSDEQEVMTWI-YYDGLQAGSKVKEIVVPIIIVNGTPVNATREV 234  
DB 817 VAYDIWTNSTPTTTG-OPNGTEIMLWNSRGVQVPGSQATGV-----TVAGHTWNV 868  
QY 235 WKA-NIGWEYVAFRIKTPKEGTVTIPYGAFISVAANISLNPNTYLYEDVEIGTE 290  
DB 869 WQOQQTWNKIISY-VLTGATISINLDKAIFADAAARGSL--NTSDYLLDVEAGFE 922

RESULT 15

US-10-441-625-22  
Sequence 22, Application US/10441625  
Publication No. US20030203467A1  
GENERAL INFORMATION:  
APPLICANT: Guafietti, Peter  
APPLICANT: Mitchinson, Colin  
APPLICANT: Phillips, Jay Ian  
TITLE OF INVENTION: Novel Variant EGIII-Like Cellulase  
FILE REFERENCE: GC631  
CURRENT APPLICATION NUMBER: US/10/441,625  
CURRENT FILING DATE: 2003-05-19  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Streptomyces lividans Celb  
US-10-441-625-22

Query Match 6.7%; Score 114.5; DB 12; Length 381;  
Best Local Similarity 27.1%; Pred. No. 0.018;  
Matches 52; Conservative 21; Mismatches 80; Indels 39; Gaps 10;

Qy	125	YPEIFYGNKPNANYATDGP	IPLESKVS	NLTDFY	LTISYKLEPKNGLP	INFAIES	WLTRE	184
Db	97	YPSVFENGCHYTNCSPGTD	---	LPVRLDTVSAAPSSISYGF	--	VDGAVYNASYDIWLDPT	150	
Qy	185	AWRTTGINSDEQEVMIWIIYD	G-L	OPAGSKVKEI	WVPII	VNGTPV	--	NATFEVW----- 235
Db	151	A-RTDGVN--QTEIMWFWNRV	GP	IQPIGSPV	-----	GTASVCGRTWEVWSGGNGS	197	
Qy	236	-----KANIGWEY-VAFRIK	TFIK	EGTVTIPY	GAFIS	VAANISSLEPNY	TELYLEDV	285
Db	198	NDVLSFVAPSAISGWSFDVMD	FVRAT	VARGLAENDW	-Y	LTSVQAGPEPWONGAGLA	VNSF	256
Qy	286	EIGTEFGTPTT	297					
Db	257	SSTVETGTPPGT	268					

Search completed: June 29, 2004, 20:19:14  
Job time : 60 secs

[The body of the page contains extremely faint, illegible text, likely bleed-through from the reverse side of the document. The text is too light to transcribe accurately.]

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 15:30:21 : Search time 3965 Seconds  
(without alignments)  
7230.187 Million cell updates/sec

Title: US-09-914-543-45  
Perfect score: 960  
Sequence: 1 atgagcaagaaaaagtctgt.....atagaactctttattctctaa 960

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	4.4	1101	29	CNS0182P
c 2	40.8	4.2	862	29	AG146121
3	40.6	4.2	1101	29	CNS00GDR
4	40.6	4.2	1201	13	EX394545

5	40.2	4.2	772	12	BI978207
6	40	4.2	560	29	CC799961
7	40	4.2	603	29	CC800184
8	40	4.2	691	29	CC800016
c 9	40	4.2	885	13	EX425603
10	39.6	4.1	579	14	CB452594
c 11	39.4	4.1	829	29	CC927838
c 12	39.4	4.1	966	13	BU519579
c 13	39	4.1	712	13	EX416727
14	39	4.1	958	29	CG975866
15	38.6	4.0	282	9	AI645051
16	38.6	4.0	416	9	AA656102
17	38.6	4.0	566	13	EX354627
18	38.4	4.0	580	10	AW626092
19	38.4	4.0	588	29	CC800123
20	38.4	4.0	599	29	CL002260
21	38.4	4.0	601	29	CC800157
22	38.4	4.0	641	29	CL002222
23	38.4	4.0	651	29	CL002262
c 24	38.4	4.0	664	29	CC800120
25	38.4	4.0	669	29	CL002221
c 26	38.4	4.0	670	29	CL002213
27	38.4	4.0	676	29	CC800124
28	38.4	4.0	698	29	CC800131
c 29	38.4	4.0	698	29	CL002246
c 30	38.4	4.0	699	29	CL002205
c 31	38.4	4.0	699	29	CL002265
c 32	38.4	4.0	703	29	CL002219
c 33	38.4	4.0	706	29	CL002227
34	38.4	4.0	729	29	CL002204
c 35	38.4	4.0	1201	13	EX361080
c 36	38.2	4.0	513	29	CC800141
c 37	38.2	4.0	1191	9	AL558073
c 38	38	4.0	683	29	CC800140
c 39	38	4.0	742	12	BI434962
c 40	38	4.0	807	14	CK139957
41	37.8	3.9	704	14	CF806131
c 42	37.6	3.9	407	14	CD802400
c 43	37.6	3.9	517	28	BH376219
44	37.6	3.9	675	28	BH265911
45	37.6	3.9	1177	13	EX359598

## ALIGNMENTS

RESULT 1  
CNS0182P  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CNS0182P 1101 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence Sp6 end of BAC  
BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
AL108811  
AL108811.1 GI:5629115  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CERH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC Project Grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector





```

/clone.lib="Old Blush petal SMART library"
/notes="Organ: Petal; Vector: pTriplex2; Site_1: Sfil;
Site_2: Sfil"

ORIGIN
Query Match      4.2%; Score 40.2; DB 12; Length 772;
Best Local Similarity 49.3%; Pred. No. 7.2;
Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 587 AAGAGTAATGATATGAGTTTACTATGACGATTACAAACGGCTGGCTCCAAAGTTAAGG 646
      |||
Db 340 AAGGAGATTGGTTGCAATTAACATATGATGCCGTGCGACAAATGGTATTTGTTGATA 399
      |||
QY 647 AGATTGTAGTCCCAATAATAGTTAACGGAACACAGTAATGCTACATTTGAAGTATGGA 706
      |||
Db 400 GCACGTGTTGATCAATTCATGAGAAAGCACACCTGTGATCTTCCTTTGGACGAAATC 459
      |||
QY 707 AGGCAACATTTGGTTGGAGTATGTCATTTAGTAATAAGACCCCAATCAAGAGGGAA 766
      |||
Db 460 AGATTATAAAGGCTTGGAGGAGGTTCTGTTGGCATGAGAAATCGAGGAAAGAGAGAG 519
      |||
QY 767 CAGTCACAATTCATACGAGCATTTTAAAGT 799
      |||
Db 520 CATTGATACCTCCTCTGTGGATACACAATG 552
      |||

RESULT 6
LOCUS CC799961
DEFINITION 0250069-08A1-A09 UniformMu MutAIL Library Zea mays genomic clone
ACCESSION CC799961
VERSION CC799961.1 GI:32622384
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 560)
AUTHORS Latshaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize
population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
0250069-08, Primer set: A
Class: transposon insertion site.
Location/Qualifiers
1..560
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="0250069-08A1-A09"
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon
insertions in Mu inactive lines were extracted from the
UniformMu maize population by the thermo asymmetric
interlaced PCR (TAIL) protocol using primers specific for
the Mu terminal inverted repeat and a set of 16 arbitrary
primers. Amplicons were size enriched using Sepharose 400
spin columns and cloned into the TOPO PCR4 vector."

FEATURES
source
1..560
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="0250069-08A1-A09"
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon
insertions in Mu inactive lines were extracted from the
UniformMu maize population by the thermo asymmetric
interlaced PCR (TAIL) protocol using primers specific for
the Mu terminal inverted repeat and a set of 16 arbitrary
primers. Amplicons were size enriched using Sepharose 400
spin columns and cloned into the TOPO PCR4 vector."

ORIGIN
Query Match      4.2%; Score 40; DB 29; Length 560;
Best Local Similarity 53.1%; Pred. No. 7.1;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 786 AGCATTTTAAAGTTGTGAGCCCAACATTTCAAGCTTACCAATTTACACAGAACTTACTT 845
      |||
Db 378 AACATTTTAAACGACACACTGAAAATTTACAACTTTTATTAATGGCTCACAACATTGCTT 437
      |||
QY 846 AGAGGACGTGGAGATGGAACTGAGTTTGGACGCCAGCACTACTCCGCCACCTAGA 905
      |||
Db 438 ATAAGAGGTACAGAGTATAAAACAGAAATGCTTGAAGAGGTACTCCCTCCGTTCTTTTA 497
      |||
QY 906 GTGGTGGATCAAAAACATAACACTAACTCCTCTAGATAGA 945
      |||

```

```

Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 786 AGCATTTTAAAGTTGTGAGCCCAACATTTCAAGCTTACCAATTTACACAGAACTTACTT 845
      |||
Db 376 AACATTTTAAACGACACACTGAAAATTTACAACTTTTATTAATGGCTCACAACATTGCTT 435
      |||
QY 846 AGAGGACGTGGAGATGGAACTGAGTTTGGACGCCAGCACTACTCCGCCACCTAGA 905
      |||
Db 436 ATACGAGGTACAGAGTATCAAAACAGAAATGCTTGAAGAGGTACTCCCTCCGTTCTTGCGA 495
      |||
QY 906 GTGGTGGATCAAAAACATAACACTAACTCCTCTAGATAGA 945
      |||
Db 496 TTTGTCGTGATAGCGTAAATTCACACTATCCAGCGACA 535
      |||

RESULT 7
LOCUS CC800184
DEFINITION 02S0069-08D1-C08 UniformMu MutAIL Library Zea mays genomic clone
ACCESSION CC800184
VERSION CC800184.1 GI:32622727
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 603)
AUTHORS Latshaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize
population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
02S0069-08, Primer set: D
Class: transposon insertion site.
Location/Qualifiers
1..603
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0069-08D1-G08"
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon
insertions in Mu inactive lines were extracted from the
UniformMu maize population by the thermo asymmetric
interlaced PCR (TAIL) protocol using primers specific for
the Mu terminal inverted repeat and a set of 16 arbitrary
primers. Amplicons were size enriched using Sepharose 400
spin columns and cloned into the TOPO PCR4 vector."

ORIGIN
Query Match      4.2%; Score 40; DB 29; Length 603;
Best Local Similarity 53.1%; Pred. No. 7.3; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 786 AGCATTTTAAAGTTGTGAGCCCAACATTTCAAGCTTACCAATTTACACAGAACTTACTT 845
      |||
Db 378 AACATTTTAAACGACACACTGAAAATTTACAACTTTTATTAATGGCTCACAACATTGCTT 437
      |||
QY 846 AGAGGACGTGGAGATGGAACTGAGTTTGGACGCCAGCACTACTCCGCCACCTAGA 905
      |||
Db 438 ATAAGAGGTACAGAGTATAAAACAGAAATGCTTGAAGAGGTACTCCCTCCGTTCTTTTA 497
      |||
QY 906 GTGGTGGATCAAAAACATAACACTAACTCCTCTAGATAGA 945
      |||

```





```

RESULT 10
LOCUS      CB452594
DEFINITION 707465 MARC 6BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  CB452594
VERSION     CB452594.1
KEYWORDS   EST, 29258976
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus

REFERENCE
AUTHORS    Smith, T.P.L., Roberts, A.J., Echtenkamp, S.E., Chitko-McKown, C.G.,
            Wray, J.E. and Keele, J.W.
TITLE      A second set of bovine ESTs from pooled-tissue normalized libraries
JOURNAL    Unpublished (2003)
COMMENT    USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@mail.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: FQY8064 row: C column: 13
            Seq primer: GTAATACGACTCACTATAGG.
FEATURES   source
            Location/Qualifiers
            1..579
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /clone_lib="MARC 6BOV"
            /note="vector: pcDNA3.1, Site 1: EcoRI; Site 2: NotI;
            Library made with RNA pooled from multiple tissues
            including liver, lung, hypothalamus, pituitary, and
            placenta/endometrium."
ORIGIN
Query Match      4.1%; Score 39.6; DB 14; Length 579;
Best Local Similarity 54.9%; Pred. No. 9.3;
Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 694 TTGAAGTATGGAAGCAACATTTGGTGGAGATGTTGCATTTAGAATAAGACCCCA 753
DB 297 TTTTAATTCCTCAATCTTACTTCTGTGGAATATAAGGGTTTGAACTTTAATACAAG 356

QY 754 ATCAAGAGGGAAGTGCACATTCATCAGGACATTTATAGTGTTCGAGCCAACTT 813
DB 357 TTAATAAATTTAACAATTAATTTTCAATGACCATTTTAAGTCCTTAAGACAAAAT 416

QY 814 TCAAGCTTACCAATTTACACAG 835
DB 417 CTTAACTTTCTATTGCACAG 438

RESULT 11
LOCUS      CC927838/c
DEFINITION CC927838
ACCESSION  CC927838
VERSION     CC927838.1
KEYWORDS   GSS, 33578936
SOURCE     Zea mays subsp. mays (maize)
ORGANISM   Zea mays subsp. mays

REFERENCE
AUTHORS    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 829)

```

```

AUTHORS      Bharti, A.K., Young, S., Kavchok, S., Keizer, G., Bronzino, A.C.,
            Rouzard, K., Fuks, G., Yu, Y., Wing, R. and Messing, J.
TITLE        Sequencing of the maize genome at PGIR (2003b)
JOURNAL      Unpublished (2003)
COMMENT      Contact: Bharti, A.K.
            Dr. Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 102.
FEATURES     source
            Location/Qualifiers
            1..829
            /organism="Zea mays subsp. mays"
            /mol_type="genomic DNA"
            /cultivar="B73"
            /sub_species="mays"
            /db_xref="taxon:4578"
            /clone="ZMMBB0530M15"
            /lab_host="E. coli DH10B"
            /clone_lib="ZMMBBc"
            /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"
ORIGIN
Query Match      4.1%; Score 39.4; DB 29; Length 829;
Best Local Similarity 53.6%; Pred. No. 12;
Matches 82; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 661 ATAATAGTTACGGAACACCACTAAATGCTCATTTGAAGTATGGAAGCAAACTTGGT 720
DB 584 ATTTAGAGACTAAATAAATTTATGTCATCAATAAATGTTGGATGTCAAATATAAGT 525

QY 721 TGGGAGTATGTCATTTAGATAAGACCCCAATCAAGAGGGAACAGTCACCAATTCGA 780
DB 524 ACTAATAATAGTCTAATTTATTAATAAATTAATTCATCGATTAAGACTAAAAGACATTTT 465

QY 781 TACGAGCATTTATTAAGTGTTCAGCCCAACT 813
DB 464 TAAGAATAATTAATTTGTGGAGGCAAAATATT 432

RESULT 12
LOCUS      BUS19579/c
DEFINITION AGENCOURT 10166786 NIH MGC 134 Mus musculus cDNA clone
ACCESSION  BUS19579
VERSION     BUS19579.1
KEYWORDS   EST, 22827105
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 966)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
            CDNA Library Preparation: Invitrogen Corp
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC Clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILLNL at:
            http://image.llnl.gov
            Plate: LLAM14097 row: i column: 08
            High quality sequence start: 62
            High quality sequence stop: 563.

```



Search completed: July 2, 2004, 00:46:15  
Job time : 3969 secs

QY 64 TTGTGAGAAAGTATCATACCTCTGAGGACCAAGTCAACTTCAAAATACCTCA 114  
|||||  
Db 469 TTGTGATATATCTAATGCGTATATATAGAGTATTATATATATATATATCA 519

RESULT 15  
AI645051 282 bp mRNA linear EST 29-APR-1999  
LOCUS vs46e05.v1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone  
DEFINITION IMAGE:1149344 5', mRNA sequence.

ACCESSION AI645051  
VERSION AI645051.1 GI:4723526  
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 282)  
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R., and Wilson, R.  
The WashU-NCI Mouse EST Project 1999

REFERENCE Unpublished (1999)

JOURNAL Contact: Marra M/WashU-NCI Mouse EST Project 1999

COMMENT Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:622552

This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)

Seq primer: -40RP from Gibco

High quality sequence stop: 273.

Location/Qualifiers

FEATURES

source  
1..282  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1149344"  
/tissue\_type="Tcell"  
/dev\_stage="M30 CD4+ cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse Tcell 937311"  
/note="Organ: blood; Vector: pBluescript SK-; Site 1:  
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;  
Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGCAGG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

ORIGIN

Query Match 4.0%; Score 38.6; DB 9; Length 282;  
Best Local Similarity 56.8%; Pred No. 13;  
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 25 GTATCTATCTTAACAACTCTTTTAGTACAGCAATATTTTGTAGAAAGTATCATACC 84  
Db |||||  
57 GAATGTATAAAAGTATCTTTGTGTAAAGCAACCAATTTGTGAAGCAGACCAATACA 116  
QY 85 TCTGGAGCAAGTCAACTCAATACCTCTACACCCCAACCAACACACTTCCACT 144  
Db |||||  
117 CACATGACATATCATCTTCCTCCCAATGACAGCTCCCTCTGTAGACACACTGAAGT 176

QY 145 ACCAA 149

Db |||||

177 AGCAA 181

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 19:53:43 ; Search time 61 Seconds  
(without alignments)  
1477.584 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701  
Sequence: 1 MSXXKFVIVSILRILLVQAI.....HLEWITNITLTPLDRLPLIS 319

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	2	AAW35007 Pyrococcus
2	1701	100.0	319	2	AAW29729 Pyrococcus
3	1701	100.0	319	2	AAW49875 Pyrococcus
4	1701	100.0	319	2	AAW23155 Pyrococcus
5	1701	100.0	319	5	AAW50997 Pyrococcus
6	1701	100.0	319	5	AAW50997 Pyrococcus
7	1701	100.0	319	7	AAW50997 Pyrococcus
8	1617	95.1	300	6	ADP96635 Pyrococcus
9	486.5	28.6	288	2	AAW23156 Dictyoglo
10	480	28.2	297	2	AAW57777 Recombina
11	480	28.2	297	2	AAW29730 Endo-beta
12	439.5	25.8	332	2	AAW17865 Sulfolobu
13	430	25.3	332	2	AAW17866 Sulfolobu
14	328	19.3	841	2	AAW34985 Archaeabac
15	166.5	9.8	243	5	AAW50977 Truncated
16	166.5	9.8	244	5	AAW50975 Truncated
17	166.5	9.8	261	5	AAW50975 Rhodother
18	166.5	9.8	261	5	AAW50975 Rhodother
19	165.5	9.7	240	5	AAW50980 Truncated
20	165.5	9.7	241	5	AAW50979 Truncated
21	165.5	9.7	242	5	AAW50978 Truncated
22	165.5	9.7	628	2	AAW34999 Archaeabac
23	164.5	9.7	239	5	AAW50981 Truncated
24	163.5	9.6	233	5	AAW50987 Truncated
25	163.5	9.6	234	5	AAW50986 Truncated

26	163.5	9.6	235	5	AAW50985 Truncated
27	163.5	9.6	236	5	AAW50984 Truncated
28	163.5	9.6	237	5	AAW50983 Truncated
29	163.5	9.6	238	5	AAW50982 Truncated
30	160.5	9.4	224	5	AAW50996 Truncated
31	160.5	9.4	225	5	AAW50995 Truncated
32	160.5	9.4	226	5	AAW50994 Truncated
33	160.5	9.4	227	5	AAW50993 Truncated
34	160.5	9.4	228	5	AAW50992 Truncated
35	160.5	9.4	229	5	AAW50991 Truncated
36	160.5	9.4	230	5	AAW50990 Truncated
37	160.5	9.4	231	5	AAW50989 Truncated
38	160.5	9.4	232	5	AAW50988 Truncated
39	143.5	8.4	261	2	AAW88462 Bacillus
40	132.5	7.8	429	3	AAW84346 Amino aci
41	128	7.5	260	2	AAW06369 Rhodother
42	128	7.5	260	3	AAW84347 Amino aci
43	128	7.5	260	3	AAW14882 Emericell
44	128	7.5	260	3	AAU77590 R. marinu
45	128	7.5	260	5	AAU77434 Rhodother

## ALIGNMENTS

### RESULT 1

AAW35007  
ID AAW35007 standard; protein; 319 AA.  
XX  
AC AAW35007;  
AC  
DT 17-OCT-2003 (revised)  
DT 21-MAY-1998 (first entry)  
XX  
DE Pyrococcus furiosus endoglucanase.  
XX  
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;  
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;  
KW thermostable enzyme; thermophilic; glycosidase.  
XX  
OS Pyrococcus furiosus; (Clone 7EG1).  
XX  
PN WO9744361-A1.  
XX  
PD 27-NOV-1997.  
XX  
PF 22-MAY-1997; 97WO-US008793.  
XX  
PR 22-MAY-1996; 96US-00651572.  
XX  
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.  
XX  
PI Lam DE, Mathur BJ;  
XX  
DR WPI; 1998-018435/02.  
XX  
N-PSDB; AAT94215.  
XX  
PT Endo:glucanase(s), preferably form archaeal bacterium, ABPII 1a - useful  
PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic  
PT bonds in cellulose.  
XX  
PS Claim 1; Fig 1W; 164pp; English.  
XX  
CC This protein comprises an endoglucanase of Pyrococcus furiosus (Clone  
CC 7EG1), that is capable of degrading carboxymethylcellulose and of  
CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has homology  
CC to an endoglucanase of archaeobacterium ABPIIa (see AAW34983). It can be  
CC produced from native cells or from recombinant host cells, especially  
CC prokaryotic host cells transformed with a plasmid or virus-derived vector  
CC including the endoglucanase DNA (see AAT94215). 24 Endoglucanases (see  
CC AAW34986-W35008) are claimed. They can be used to degrade cellulose for  
CC the conversion of plant biomass into fuels and chemicals, for use in  
CC detergents, textiles, animal feed, waste treatment, and in the fruit

CC juice and brewing industries for the clarification and extraction of  
 CC juices. (Updated on 17-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 2; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-150;  
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKKKEVIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60  
 Db 1 MSKKKEVIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60

Qy 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120  
 Db 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120

Qy 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
 Db 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

Qy 181 LTREAWRTTGINSDEQEWMIWYDGLQAGSKVKEIWPVIIVNGTVPVATFEVWKANIG 240  
 Db 181 LTREAWRTTGINSDEQEWMIWYDGLQAGSKVKEIWPVIIVNGTVPVATFEVWKANIG 240

Qy 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNYTELYLEDVEIGTFGTPSTTSAH 300  
 Db 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNYTELYLEDVEIGTFGTPSTTSAH 300

Qy 301 LEWMITNITLTPLDRPLIS 319  
 Db 301 LEWMITNITLTPLDRPLIS 319

RESULT 2  
 AAW29729  
 ID AAW29729 standard; protein; 319 AA.  
 XX  
 AC AAW29729;  
 XX  
 DT 18-NOV-1998 (first entry)  
 DE Pyrococcus furiosus endo-beta-1, 4-glucanase.  
 XX  
 KW Pyrococcus furiosus endo-beta-1, 4-glucanase; cellulolytic activity;  
 KW textile industry; cellulosic fibre; in industrial cleaning process;  
 KW sugar; instant coffee; oil industry; hydrocolloid cellulose derivative;  
 KW drilling.  
 XX  
 OS Pyrococcus furiosus.  
 XX  
 PN WO9833895-A1.  
 XX  
 PD 06-AUG-1998.  
 XX  
 PF 30-JAN-1998; 98WO-DK000039.  
 XX  
 PR 31-JAN-1997; 97DK-00000114.  
 PR 11-JUL-1997; 97DK-00000853.  
 XX  
 XX (NOVO ) NOVO-NORDISK AS.  
 PA  
 XX Andersen L, Bjornvad MB, Schuelein M;  
 PI  
 XX WPI; 1998-437450/37.  
 DR  
 DR N-PSDB; AAV47540.  
 XX  
 XX Isolated endo-beta-1,4-glucanase - used for e.g. treating cellulosic  
 PT fibres or polymers, feed production or in oil industry for enhancing oil  
 PT recovery.  
 XX  
 XX Disclosure; Page 45-46; 56pp; English.  
 PS  
 XX

CC The present sequence represents the Pyrococcus furiosus endo-beta-1, 4-  
 CC glucanase (EG) protein sequence. The invention provides for an enzyme  
 CC composition having EG activity which has optimum activity at a  
 CC temperature of at least 90 deg. C. The EG enzyme exhibit cellulolytic  
 CC activity at extremely high temperatures in a very broad pH range.  
 CC Therefore it is claimed to be useful for, e.g. in the textile industry  
 CC for improving the properties of cellulosic fibres or fabric; for  
 CC providing a stone-washed look of denim; in industrial cleaning processes;  
 CC in the conversion of biomass to sugars; in the production of instant  
 CC coffee or similar extraction processes or in the oil industry for  
 CC degradation of aqueous solutions of hydrocolloid cellulose derivatives  
 CC used in drilling  
 XX  
 SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 2; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-150;  
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKKKEVIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60  
 Db 1 MSKKKEVIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60

Qy 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120  
 Db 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120

Qy 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
 Db 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

Qy 181 LTREAWRTTGINSDEQEWMIWYDGLQAGSKVKEIWPVIIVNGTVPVATFEVWKANIG 240  
 Db 181 LTREAWRTTGINSDEQEWMIWYDGLQAGSKVKEIWPVIIVNGTVPVATFEVWKANIG 240

Qy 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNYTELYLEDVEIGTFGTPSTTSAH 300  
 Db 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNYTELYLEDVEIGTFGTPSTTSAH 300

Qy 301 LEWMITNITLTPLDRPLIS 319  
 Db 301 LEWMITNITLTPLDRPLIS 319

RESULT 3  
 AAW49875  
 ID AAW49875 standard; protein; 319 AA.  
 XX  
 AC AAW49875;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 21-DEC-1998 (first entry)  
 XX  
 XX Pyrococcus VC1-7EG1 glycosidase.  
 DE  
 XX  
 KW Glycosidase; VC1-7EG1; thermostable enzyme; oligosaccharide; glucose;  
 KW sugar; baking; textile; detergent.  
 XX  
 OS Pyrococcus furiosus; strain VC1-7EG1.  
 XX  
 PN WO9824799-A1.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PF 08-DEC-1997; 97WO-US022623.  
 PR  
 PR 06-DEC-1996; 96US-0056916P.  
 PR 10-OCT-1997; 97US-00949026.  
 XX  
 PA (DIVE-) DIVERSA CORP.  
 XX  
 XX Bylina EJ, Swanson RV, Mathur EJ, Lam DE;  
 XX

DR WPI: 1998-362407/31.  
 DR N-PSDB; AAV36924.  
 XX  
 PT Glycosidase enzymes from organisms of the genera Staphylothermus,  
 PT Pyrococcus and Thermococcus - for deriving sugar from oligosaccharides,  
 PT useful in the e.g. food processing, textile or baking industries.  
 XX  
 PS Claim 1; Fig 18a-b; 92pp; English.  
 XX  
 CC This is the amino acid sequence of glycosidase VCI-75G1, deduced from a  
 CC polynucleotide (see AAV36924) of a clone (75G1) of Pyrococcus furiosus  
 CC VCI, which grows optimally at 100 degC. The invention provides 18  
 CC polynucleotides (see AAV36907-24) coding for thermostable glycosidases  
 CC (see AAV49858-75) having glycosidase, alpha-galactosidase, beta-  
 CC galactosidase, beta-mannosidase, beta-mannanase, endoglucanase or  
 CC pullulanase activity. Vectors and host cells are also claimed. A method  
 CC is provided for producing the enzymes by recombinant techniques. A  
 CC claimed method for generating glucose from soluble cell oligosaccharides  
 CC comprises contacting a sample (selected from dairy products, fruit juice,  
 CC detergent, textile, guar gum, animal feed, plant biomass or waste  
 CC product) containing oligosaccharides (selected from maltose, cellobiose,  
 CC lactose, sucrose, raffinose, stachyose, verbascone, cellulose, starch,  
 CC amylose, glycogen, disaccharides, polysaccharides and pullulan) with one  
 CC of the claimed glycosidases such that glucose is produced. (Updated on 17  
 CC -Oct-2003 to standardise OS field)  
 XX  
 SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 2; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-150;  
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSKKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSTSTPPQTLSTTKVLRYPDDG 60  
 DB 1 MSKKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSTSTPPQTLSTTKVLRYPDDG 60  
 QY 61 EWPAPIDKDGNGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120  
 DB 61 EWPAPIDKDGNGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120  
 QY 121 WHGYPEIFYGNKPWNANYATDGPILPSPKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
 DB 121 WHGYPEIFYGNKPWNANYATDGPILPSPKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
 QY 181 LTREAWRTTGINSDEQVMWIIYDGLQAGSKVKEIWPVPIIIVNGTVPVNAFVWKANIG 240  
 DB 181 LTREAWRTTGINSDEQVMWIIYDGLQAGSKVKEIWPVPIIIVNGTVPVNAFVWKANIG 240  
 QY 241 WEYVAFRIKTPIKEGTVTIPYGAFISVAANISSLPNTYLEDVEIGTFEFGTPTSTSAH 300  
 DB 241 WEYVAFRIKTPIKEGTVTIPYGAFISVAANISSLPNTYLEDVEIGTFEFGTPTSTSAH 300  
 QY 301 LEWWTNITLTPLDRPLIS 319  
 DB 301 LEWWTNITLTPLDRPLIS 319

RESULT 4  
 AAY23155  
 ID AAY23155 standard; protein; 319 AA.  
 AC AAY23155;  
 XX  
 XX 24-AUG-1999 (first entry)  
 DT  
 XX Pyrococcus cellulase protein sequence.  
 DE Cellulose containing fabric; cellulase; fabric handle; fabric appearance;  
 XX pilling resistance.  
 KW Pyrococcus sp.  
 OS Pyrococcus sp.  
 XX WO9932708-A1.  
 PN

XX 01-JUL-1999.  
 PD 17-DEC-1998; 98WO-US026798.  
 XX 19-DEC-1997; 97US-0068274P.  
 PR  
 XX (NOVO ) NOVO NORDISK BIOCHEM NORTH AMERICA.  
 PA  
 XX Liu J, Condon B;  
 PI WPI: 1999-395422/33.  
 DR  
 XX Treatment of cellulose containing fabric using cellulase, achieves better  
 PT handling, appearance and pilling resistance.  
 PT  
 XX Claim 7; Page 25; 31pp; English.  
 PS  
 XX The specification describes a method for treating a cellulose containing  
 CC fabric, comprising contacting the fabric with an aqueous bulk solution  
 CC containing a cellulase, and subjecting the contacted fabric to high  
 CC temperature. The method of treatment is useful for treating cellulose-  
 CC containing fabrics. The method of the invention achieves better fabric  
 CC handle, appearance and pilling resistance. The quality of the cellulosic  
 CC fabric is enhanced and allows uniform action by the cellulase. The  
 CC present sequence represents a cellulase which is used in the method of  
 CC the invention.  
 CC  
 XX  
 SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 2; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-150;  
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSKKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSTSTPPQTLSTTKVLRYPDDG 60  
 DB 1 MSKKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSTSTPPQTLSTTKVLRYPDDG 60  
 QY 61 EWPAPIDKDGNGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120  
 DB 61 EWPAPIDKDGNGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120  
 QY 121 WHGYPEIFYGNKPWNANYATDGPILPSPKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
 DB 121 WHGYPEIFYGNKPWNANYATDGPILPSPKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
 QY 181 LTREAWRTTGINSDEQVMWIIYDGLQAGSKVKEIWPVPIIIVNGTVPVNAFVWKANIG 240  
 DB 181 LTREAWRTTGINSDEQVMWIIYDGLQAGSKVKEIWPVPIIIVNGTVPVNAFVWKANIG 240  
 QY 241 WEYVAFRIKTPIKEGTVTIPYGAFISVAANISSLPNTYLEDVEIGTFEFGTPTSTSAH 300  
 DB 241 WEYVAFRIKTPIKEGTVTIPYGAFISVAANISSLPNTYLEDVEIGTFEFGTPTSTSAH 300  
 QY 301 LEWWTNITLTPLDRPLIS 319  
 DB 301 LEWWTNITLTPLDRPLIS 319

RESULT 5  
 AAM50997  
 ID AAM50997 standard; protein; 319 AA.  
 AC AAM50997;  
 XX  
 XX 15-MAY-2002 (first entry)  
 DT  
 XX Pyrococcus furiosus thermostable cellulase EglA.  
 DE Cellulase; EglA; thermostable; enzyme.  
 KW Pyrococcus furiosus.  
 OS Pyrococcus furiosus.  
 XX

PH	Key	Location/Qualifiers	
FT	Peptide	1..19	
FT	Protein	/label= Signal_peptide	
FT	Protein	20..319	
FT	Domain	/label= Mature_protein	
FT	Domain	28..49	
FT	Domain	/label= Linker_moiety	
FT	Domain	50..319	
FT	Domain	/label= Catalytic_domain	
XX	WO200196382-A2.		
PN	20-DEC-2001.		
XX	15-JUN-2001; 2001WO-IS000012.		
XX	15-JUN-2000; 2000US-00594894.		
XX	(PROK-) PROKARIA EHF.		
XX	Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;		
XX	WPI; 2002-226850/28.		
XX	N-PSDB; ABA91905.		
XX	Novel thermostable variant cellulase which is truncated such that amino		
XX	terminal hydrophobic region and linker group of corresponding full length		
XX	enzyme is deleted, has improved catalytic properties and/or stability.		
XX	Disclosure; Page 48-49; 51pp; English.		
XX	The present sequence is that of the Pyrococcus furiosus thermostable		
XX	cellulase EglA, a family 12 glycosyl hydrolase. The invention provides		
XX	polypeptides having thermostable cellulase activity. These are variants		
XX	of a glycosyl hydrolase of family 12 and are truncated such that 1 or		
XX	more of the amino acid residues corresponding to position 1 to about 40		
XX	are deleted. The polypeptides are particularly derived from thermophilic		
XX	Rhodothermus and Pyrococcus spp. For EglA, amino acids from the N-		
XX	terminal hydrophobic region and/or the linking moiety are deleted.		
XX	Claimed polypeptides (see AAM50976-996) are also derived from		
XX	Rhodothermus marinus Cell12A cellulase. Isolated nucleic acids encoding		
XX	these thermostable cellulase polypeptides, and host cells, are also		
XX	claimed. The thermostable variant cellulases are useful e.g. in the wood		
XX	and paper pulp industries, in detergent compositions, in the textile		
XX	industry, to improve the feed value of animal feedstuffs, and in the food		
XX	industry		
XX	Sequence 319 AA;		
SQ	Query Match	100.0%; Score 1701; DB 5; Length 319;	
	Best Local Similarity	100.0%; Pred. No. 3.1e-150;	
	Matches 319; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSTSTPPQTTLSKTKIRYPDDG	60
DB	1	MSKKFVIVSILTLVQAIYFVEKYHTSEDKSTNTSTSTPPQTTLSKTKIRYPDDG	60
QY	61	EWFGAPIDKDGNGNPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQOLDNIVLRDSN	120
DB	61	EWFGAPIDKDGNGNPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQOLDNIVLRDSN	120
QY	121	WHGYPEIFYGNKPNVANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW	180
DB	121	WHGYPEIFYGNKPNVANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW	180
QY	181	LTRAWRTTGINSDEQEWIMTYIDGLOPAGSKVEIIVPIVINGTPVNAVFEVWKANIG	240
DB	181	LTRAWRTTGINSDEQEWIMTYIDGLOPAGSKVEIIVPIVINGTPVNAVFEVWKANIG	240
QY	241	WEYVAFRIKTPKEGTVITPYGAFISVAANISSLPNYELVEDVEIGTEFGTSTSAH	300
DB	241	WEYVAFRIKTPKEGTVITPYGAFISVAANISSLPNYELVEDVEIGTEFGTSTSAH	300

QY	301	LEWWTNITLTPLDRPLIS	319
DB	301	LEWWTNITLTPLDRPLIS	319
XX	RESULT 6		
XX	AAG79624		
ID	AAG79624	standard; protein; 319 AA.	
XX	AC	AAG79624;	
XX	DT	07-FEB-2003 (first entry)	
XX	DE	P. furiosus EglA.	
XX	XX	Variant; thermostable; cellulase; Cell12A; family 12; EglA; enzyme;	
XX	XX	glycosyl hydrolase; freeness; ink; coating; toner; colour; wood;	
XX	XX	paper pulp; detergent; cellulose-containing textile; garment; lint;	
XX	XX	fibrous crop; fruit; vegetable; grain; feed value; stability; solubility;	
XX	XX	catalytic activity; cytotoxicity.	
XX	OS	Pyrococcus furiosus.	
XX	XX	Key	Location/Qualifiers
XX	XX	Peptide	1..19
XX	XX	Peptide	/label= Hydrophobic peptide
XX	XX	Peptide	28..49
XX	XX	Peptide	/label= Linker peptide
XX	XX	Domain	50..319
XX	XX	Domain	/label= Catalytic domain
XX	XX	US2002102699-A1.	
XX	XX	01-AUG-2002.	
XX	XX	23-OCT-2001; 2001US-00003759.	
XX	XX	15-JUN-2000; 2000US-00594884.	
XX	XX	15-JUN-2001; 2001WO-IS000012.	
XX	XX	(PROK-) PROKARIA LTD.	
XX	XX	Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;	
XX	XX	WPI; 2002-749592/81.	
XX	XX	N-PSDB; ABA00497.	
XX	XX	Novel isolated nucleic acid encoding a polypeptide having thermostable	
XX	XX	cellulase activity, useful for producing thermostable cellulase	
XX	XX	polypeptide and as probes for isolating homologous sequences.	
XX	XX	Disclosure; Page 15-16; 21pp; English.	
XX	XX	This sequence shows EglA, a variant family 12 glycosyl hydrolase derived	
XX	XX	from P. furiosus. The amino terminal of EglA, amino acids 1-19,	
XX	XX	constitute the hydrophobic domain. Residues 28-49 constitute the linker	
XX	XX	moiety, with amino acids 50-319 forming the catalytic domain. EglA is	
XX	XX	useful to improve freeness and to remove inks, coatings, toners and	
XX	XX	colours from wood or paper pulp, in detergent compositions and to treat	
XX	XX	cellulose-containing textiles and garments to improve the feel of the	
XX	XX	fabric or to remove lint, in the treatment of fibrous crops, fruits and	
XX	XX	vegetables or grains to improve feed values or to extract starches (e.g.	
XX	XX	sugars) or other components of the crop, fruit, vegetable or grain being	
XX	XX	treated. The variant glycosyl hydrolase polypeptide has improved	
XX	XX	characteristics, such as increased solubility in aqueous solvents, increased	
XX	XX	detergent stability, increased stability (e.g. thermal stability,	
XX	XX	catalytic activity, catalytic rate) and/or	
XX	XX	reduced cytotoxicity relative to the native or full-length thermostable	
XX	XX	cellulase, but retains the substrate specificity of the native or full-	
XX	XX	length cellulase	
XX	XX	Sequence 319 AA;	



Query Match 100.0%; Score 1701; DB 5; Length 319;  
Best Local Similarity 100.0%; Pred. No. 3.1e-150;  
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKKFVIVSILITILLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60  
DB 1 MSKKKFVIVSILITILLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60

QY 61 EWPAGAPDKDGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSN 120  
DB 61 EWPAGAPDKDGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSN 120

QY 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180  
DB 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180

QY 181 LTREAWRTTGINSDEQVMIWIYDGLQAGSKVKEIIVPFIIVNGTVPVNATFEVWKANIG 240  
DB 181 LTREAWRTTGINSDEQVMIWIYDGLQAGSKVKEIIVPFIIVNGTVPVNATFEVWKANIG 240

QY 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300  
DB 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300

QY 301 LEWMITNITLPLDRPLIS 319  
DB 301 LEWMITNITLPLDRPLIS 319

RESULT 7  
ADC26955  
ID ADC26955 standard; protein; 319 AA.  
AC ADC26955;  
XX  
DT 18-DEC-2003 (first entry)  
DE Pyrococcus furiosus glycosidase enzyme SEQ ID NO:60.  
KW enzyme delivery matrix; thermostable enzyme; glycosidase; hydrolysis;  
KW lactose; galactose; glucose; guar gum; animal feed; hydraulic fracturing;  
KW oil recovery; gas recovery; biodegradable; enzyme.  
XX  
OS Pyrococcus furiosus.  
XX  
PN WO2003072717-A2.  
XX  
PD 04-SEP-2003.  
XX  
PF 21-FEB-2003; 2003WO-US0005189.  
XX  
PR 21-FEB-2002; 2002US-00081475.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Miller CA;  
XX  
XX WPI; 2003-748224/70.  
DR N-PSDB; ADC26951.  
XX  
PT Enzyme delivery matrix used as adjuvant for animal feeds comprises  
PT discrete particles each comprising grain germ that is spent of oil and  
PT thermostable enzyme.  
XX  
XX Disclosure; SEQ ID NO 64; 103pp; English.  
XX  
CC The present invention describes an enzyme delivery matrix (I) comprising  
CC discrete particles each comprising grain germ that is spent of oil and a  
CC thermostable enzyme, where the particle readily disperses the glycosidase  
CC enzyme into aqueous media. Also described: (1) a method for preparing an  
CC enzyme delivery matrix; (2) isolated nucleic acid molecules encoding the  
CC enzymes; (3) utilising enzymes for hydrolysing lactose to galactose and  
CC glucose; (4) utilising enzymes for hydrolysing guar gum; (5) nucleic acid

CC probes; and (6) utilising enzymes or polynucleotides to generate probes.  
CC (I) can be used as an adjuvant for animal feeds, and for hydraulic  
CC fracturing in oil and gas recovery. (I) is biodegradable. The present  
CC sequence represents a glycosidase enzyme, which is used in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 7; Length 319;  
Best Local Similarity 100.0%; Pred. No. 3.1e-150;  
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKKFVIVSILITILLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60  
DB 1 MSKKKFVIVSILITILLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60

QY 61 EWPAGAPDKDGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSN 120  
DB 61 EWPAGAPDKDGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSN 120

QY 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180  
DB 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180

QY 181 LTREAWRTTGINSDEQVMIWIYDGLQAGSKVKEIIVPFIIVNGTVPVNATFEVWKANIG 240  
DB 181 LTREAWRTTGINSDEQVMIWIYDGLQAGSKVKEIIVPFIIVNGTVPVNATFEVWKANIG 240

QY 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300  
DB 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300

QY 301 LEWMITNITLPLDRPLIS 319  
DB 301 LEWMITNITLPLDRPLIS 319

RESULT 8  
ABP96635  
ID ABP96635 standard; protein; 300 AA.  
XX  
AC ABP96635;  
XX  
DT 02-JUN-2003 (first entry)  
XX  
DE Pyrococcus furiosus EGIA protein sequence SEQ ID NO:55.  
XX  
KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;  
KW pullulanase; alpha-glucosidase; glucose isomerase; glucamylase;  
KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;  
KW maltodextrin; ethanol; fermentation; beverage; enzyme.  
XX  
OS Pyrococcus furiosus.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 261 /note= "encoded by AAG"  
FT  
XX  
XX WO2003018766-A2.  
XX  
XX 06-MAR-2003.  
XX  
XX 27-AUG-2002; 2002WO-US027129.  
PF  
XX  
XX 27-AUG-2001; 2001US-0315281P.  
PR  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;  
XX WPI; 2003-268420/26.  
DR N-PSDB; ACC44577.  
XX

PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.  
 PT alpha-amylase, useful for producing plant to produce food products having  
 PT improved taste or fermentable substrates for ethanol.  
 XX Example 26; Page 113; 158pp; English.

PS The present invention describes polynucleotides which encode processing  
 CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose  
 CC isomerase, or glucoamylase) that are optimised for expression in plants.  
 CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic  
 CC processing enzymes, which are activated under suitable conditions to act  
 CC upon the desired substrate. Also described are self-processing transgenic  
 CC plants and plant parts, e.g. grain, which express one or more of these  
 CC enzymes and have an altered composition that facilitates plant and grain  
 CC processing. Also described is a method (M) for converting starch to  
 CC starch-derived products in a transformed plant part (TPP), by activating  
 CC the starch processing enzyme contained in it. Transgenic grain is useful  
 CC for preparing maltodextrin. A transformed plant (TP) can be used to  
 CC produce food products having improved taste and to produce fermentable  
 CC substrates for ethanol and fermented beverages. (M) eliminates the need  
 CC to mill or physically disrupt the integrity of plant parts prior to  
 CC recovery of starch-derived products. The present sequence represents  
 CC ECLA, which is given in the exemplification of the present invention  
 XX  
 SQ Sequence 300 AA;

Query Match 95.1%; Score 1617; DB 6; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 2e-142;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 20 IYFEKHTSDKSTNTSTPTPTTSTTKVILKIRYDDGEGWPGAPDKDGNPEFYI 79  
 DB 1 IYFEKHTSDKSTNTSTPTPTTSTTKVILKIRYDDGEGWPGAPDKDGNPEFYI 60  
 QY 80 EINHNLNATGPAETMYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPEIFYGNKPWANY 139  
 DB 61 EINHNLNATGPAETMYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPEIFYGNKPWANY 120  
 QY 140 ATDGPILPKSVNLTDFYLTISYKLEPKNGLPINFAIESWLTREAWRTTGINSDEQVM 199  
 DB 121 ATDGPILPKSVNLTDFYLTISYKLEPKNGLPINFAIESWLTREAWRTTGINSDEQVM 180  
 QY 200 IWIIYDGLQAGSKVKEIWPVPIIVNGTPVNATPEWKNIGWEYVAPRIKTIKEGVTI 259  
 DB 181 IWIIYDGLQAGSKVKEIWPVPIIVNGTPVNATPEWKNIGWEYVAPRIKTIKEGVTI 240  
 QY 260 PYGAFISVAANISSLPNTYLYEDVEIGTEFGTPSTSAHLEWITNITLTPDRPLIS 319  
 DB 241 PYGAFISVAANISSLPNTYLYEDVEIGTEFGTPSTSAHLEWITNITLTPDRPLIS 300

RESULT 9  
 AAY23156  
 ID AAY23156 standard; protein; 288 AA.  
 XX  
 AC AAY23156;  
 XX  
 DT 24-AUG-1999 (first entry)  
 XX  
 DE Dictyoglomus cellulase protein sequence.

XX Cellulose containing fabric; cellulase; fabric handle; fabric appearance;  
 XX pilling resistance.  
 XX Dictyoglomus sp.  
 XX WO9932708-A1.  
 XX 01-JUL-1999.  
 XX 17-DEC-1998; 95WO-US026798.  
 XX 19-DEC-1997; 97US-0068274P.

XX (NOVO ) NOVO NORDISK BIOCHEM NORTH AMERICA.  
 PA Liu J, Condon B;  
 PI WPI; 1999-395422/33.  
 DR Treatment of cellulose containing fabric using cellulase, achieves better  
 PT handling, appearance and pilling resistance.  
 XX Claim 7; Page 26; 31pp; English.  
 XX The specification describes a method for treating a cellulose containing  
 CC fabric, comprising contacting the fabric with an aqueous bulk solution  
 CC containing a cellulase, and subjecting the contacted fabric to high  
 CC temperature. The method of treatment is useful for treating cellulose-  
 CC containing fabrics. The method of the invention achieves better fabric  
 CC handle, appearance and pilling resistance. The quality of the cellulosic  
 CC fabric is enhanced and allows uniform action by the cellulase. The  
 CC present sequence represents a cellulase which is used in the method of  
 CC the invention  
 XX  
 SQ Sequence 288 AA;

Query Match 28.6%; Score 486.5; DB 2; Length 288;  
 Best Local Similarity 32.2%; Pred. No. 7.7e-37;  
 Matches 104; Conservative 58; Mismatches 106; Indels 55; Gaps 5;  
 QY 4 KKFVIVSILTLVQAIY-----FVEKYHTSEDKSTNTSTPTPTTSTTKVLKI 54  
 DB 2 KKSLLSLILILLITLSPSQPKYKDAFILKAPSSGDVTTKNLEPLT----- 47  
 QY 55 RYDDGEPGAPDKDGNPEFYIENLWNLNATGPAETMYNLTSGVLHYVQQLDNIV 114  
 DB 48 -----DELNFWNIANYEGNTWMAFYKEEDTVEYYADIKNIV 83  
 QY 115 LRDRSNWVHGYPEIFYGNKPWANYATDGPILPKSVNLTDFYLTISYKLEPKNGLPIN 174  
 DB 84 LKDKSNWVHGYPEIFYGYKPNAGHNSIEKLALPKVSEFPDFVFLNKNIWEKNLPIN 143  
 QY 175 FAIESWLTREAWRTTGINSDEQVMIIYYDGLQAGSKVKEIWPVPIIVNGTPVNATFEV 234  
 DB 144 FAMEWTIKPEYQKT-VTSGDIEMWVLYANRLSPAGKGVGEVKIPILNNGNQDIIWEV 202  
 QY 235 WKANIGWYVAPRIKTIKEGVTIPYGAFIG-----VAANISSL--PNYTELYLEDVEI 287  
 DB 203 YLSPMSWYVAYKSKENILQGVKIPINEFLKHLRTILANNPSRITPEKFDQMYVTWVEI 262  
 QY 288 GTEFGTPSTSAHLEWITNITL 310  
 DB 263 GTEFGDPVTTEAKFGWTFSPFDI 285

RESULT 10  
 AAW57777  
 ID AAW57777 standard; protein; 297 AA.  
 XX  
 AC AAW57777;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 26-OCT-1998 (first entry)  
 XX  
 XX Recombinant endo-1,4-beta-glucanase.

XX Endoglucanase; cellulase; thermostable enzyme; textile; biopolishing;  
 XX stone-washing; saccharification; feedstuff; coffee.  
 XX Bacillus sp.  
 XX Dictyoglomus; sp.  
 XX Chimeric.  
 XX Key Location/Qualifiers  
 FH Peptide 1..27

FT /label= Sig\_peptide  
 FT /notes= "Bacillus sp. signal peptide"  
 FT 28. .297  
 FT /label= Mat\_protein  
 FT /notes= "mature endoglucanase from Dictyoglomus sp."  
 XX WO9828410-A1.  
 XX 02-JUL-1998.  
 XX 19-DEC-1997; 97WO-DK000583.  
 XX 20-DEC-1996; 96DK-00001483.  
 XX (NOVO ) NOVO-NORDISK AS.  
 XX Schuelein M, Bjornvad ME, Norrevang IA;  
 XX WPI; 1998-377642/32.  
 XX N-PSDB; AAV40651.  
 XX New endo-1,4-beta-glucanase active at high temperature and wide pH range  
 PT - useful, e.g. for improving properties of cellulosic textiles, e.g.  
 PT reduce pilling and improve softness.  
 XX  
 XX Example 3; Page 33; 53pp; English.  
 XX  
 CC This polypeptide is encoded by an open reading frame (see AAV40651) of  
 CC plasmid pMB447A. It comprises a signal peptide from Bacillus sp. and a  
 CC novel mature thermostable endoglucanase of Dictyoglomus sp. DSM 6262. The  
 CC signal peptide portion of the construct directs the endoglucanase to the  
 CC exterior of host Bacillus sp. cells. Recombinant endoglucanase was  
 CC expressed in Bacillus subtilis. The enzyme shows optimum activity at a  
 CC temperature above 85 degC, and exhibits an activity towards CM-cellulose  
 CC at 70 degC and pH 10 that is higher than 50% relative to the activity at  
 CC 70 degC and the optimum pH. A Bacillus subtilis strain harboring an  
 CC expression plasmid encoding the thermostable endoglucanase cloned from  
 CC Dictyoglomus sp. DSM 6262 is deposited as DSM 11903. The endoglucanase  
 CC can be used to improve properties of cellulosic textiles, e.g. to reduce  
 CC pilling, improve softness, to provide a stone-washed look to denim, for  
 CC industrial cleaning, e.g. of ultrafiltration membranes and pipes, in heat  
 CC extrudable polymers (to increase degradability), in conversion of  
 CC biomass to sugars, in production of ethanol (particularly preliquefaction  
 CC of grain), for improving digestibility of fodder grains, and in  
 CC production of instant coffee or other similar extraction processes  
 CC (increasing capacity of the extraction column) (all claimed). Also  
 CC provided are DNA constructs encoding the endoglucanase, recombinant  
 CC expression vectors and host cells (preferably Bacillus, Dictyoglomus or  
 CC Saccharomyces species or a filamentous fungus). (Updated on 17-OCT-2003  
 CC to standardise OS field)  
 XX  
 XX Sequence 297 AA;  
 SQ  
 Query Match 28.2%; Score 480; DB 2; Length 297;  
 Best Local Similarity 32.1%; Pred. No. 3.3e-36;  
 Matches 105; Conservative 63; Mismatches 109; Indels 50; Gaps 7;  
 QY 1 MSKKFVIVSILTLVQALFYVEKYHTGSDKSTNSSTP-----PQTLLSTTK 50  
 DB 1 MKQOKRLYARLLTLF--ALIFLLPH-----SAAQAQTPKYKADFIKAPSSGVVTK 51  
 QY 51 VLKIRYPDDGEWFGAPIDKGDGNPFYEINLNWLNATGFAEMTYNLTSGVLHYVQQL 110  
 DB 52 NLPLT-----LELNFWNIANTEGNTWMAFYKEEDTVEYADI 88  
 QY 111 DNLVLRDRSNWVHGYPEIFYGNKPWNAVATDGPILPKSVSLNLTDFYLTISYKLEPKNG 170  
 DB 89 KNIVLKDKNWVHGYPEVYGYKFPWAGHNSIEKALPKVSEFPDVLNFKLNIWYKRN 148  
 QY 171 LPINFALESWLTREAWRTTGINSDEQEWLWIYYDGLQAGSKVKIIVPIVUNGPPVNA 230  
 DB 149 LPINFAWETWITKEPKYKT-VTSGDIEMVMVLYANRLSPAGRKVGSKVPIILNGNQDI 207

QY 231 TFEVWKANIGWYVAFRIKTIPIKEGTVTIPYGAFTS-----VAANISL--PNYTELYLE 283  
 DB 208 IWEVYLSMSPWDYVAYKSKENILOQVKIPINEFLKHLRTILANNPSRITPEKEDQMYVT 267  
 QY 284 DVEIGTEFGTSTTSIAHLEWMTITNL 310  
 DB 268 VWEIGTEFGDPTTTEAKFGWTFSPFDI 294  
 RESULT 11  
 AAW29730  
 ID AAW29730 standard; protein; 297 AA.  
 XX  
 AC AAW29730;  
 XX  
 DT 18-NOV-1998 (first entry)  
 XX  
 DE Endo-beta-1, 4-glucanase expressed from the pMB447A plasmid.  
 XX  
 DE Pyrococcus furiosus endo-beta-1, 4-glucanase; cellulolytic activity;  
 KW textile industry; cellulosic fibre; in industrial cleaning process;  
 KW sugar; instant coffee; oil industry; hydrocolloid cellulose derivative;  
 KW drilling; chimeric.  
 XX  
 OS Synthetic.  
 OS Dictyoglomus sp.  
 XX  
 PH Key Location/Qualifiers  
 FT Peptide 1..29  
 FT Protein /note= "Signal peptide derived from Bacillus species"  
 FT 30..297  
 FT /note= "Dictyoglomus species endo-beta-1, 4-glucanase"  
 XX WO9833895-A1.  
 XX 06-AUG-1998.  
 XX 30-JAN-1998; 98WO-DK000039.  
 XX 31-JAN-1997; 97DK-00000114.  
 XX 11-JUL-1997; 97DK-00000853.  
 XX (NOVO ) NOVO-NORDISK AS.  
 XX Andersen L, Bjornvad ME, Schuelein M;  
 WPI; 1998-437450/37.  
 N-PSDB; AAV47544.  
 Isolated endo-beta-1,4-glucanase - used for e.g. treating cellulosic  
 fibres or polymers, feed production or in oil industry for enhancing oil  
 recovery.  
 Disclosure; Page 32; 56pp; English.  
 The present sequence represents the Dictyoglomus species endo-beta-1, 4-  
 glucanase (EG) expressed from the pMB447A plasmid. This vector was given  
 as an example of an expression vector that could be used in the method of  
 the invention. The invention provides for a Pyrococcus furiosus endo-beta  
 -1, 4-glucanase encoding DNA (AAV47540) isolated from a P. furiosus  
 genomic DNA library. The invention provides for an enzyme composition  
 having EG activity which has optimum activity at a temperature of at  
 least 90 deg. C. The EG enzyme exhibits cellulolytic activity at  
 extremely high temperatures in a very broad pH range. Therefore it is  
 claimed to be useful for, e.g. in the textile industry for improving the  
 properties of cellulosic fibres or fabric, for providing a stone-washed  
 look of denim; in industrial cleaning processes; in the conversion of  
 biomass to sugars; in the production of instant coffee or similar  
 extraction processes or in the oil industry for degradation of aqueous  
 solutions of hydrocolloid cellulose derivatives used in drilling  
 Sequence 297 AA;  
 SQ

Query Match 28.2%; Score 480; DB 2; Length 297;  
Best Local Similarity 32.1%; Pred. No. 3.3e-36;  
Matches 105; Conservative 63; Mismatches 109; Indels 50; Gaps 7;  
QY 1 MSKKKVVIVSLTLLVQAIYFVEKYHTSBDKSTSTP-----PQTLSTTK 50  
DB 1 MKQQRLLYARLLTLF--ALIFLLPH-----SAAAAQTPKYKDAFILKAPSSGDVTK 51  
QY 51 VLKIRYDDGWEPCAPIDKDGNGPEFYIEINLWNLNATGFAEMTNLTSGVLHVYQOL 110  
DB 52 NLPT-----LELNFNFIANYEGNTWMAFYKEEDIVEYADI 88  
QY 111 DNVLRDSNVWHGYEPIFYGNKPWNAYATDGPILPISKVSNLTDPLYISYKLPKNG 170  
DB 89 KNIVLKDNKSNVWHGYEPIFYGNKPWNAYATDGPILPISKVSNLTDPLYISYKLPKNG 148  
QY 171 LPINFATESLWLTREAWRTGINSDEQEVMIWYDGLQAPGSKVKEIVVPIIVNGTPVNA 230  
DB 149 LPINFAMETWTKPEYQKT-VTSGDIEEMVWLNVANRUSPAGRKVGKIFPIILNGKQDI 207  
QY 231 TFEVVKANIGWEYVAFRIKTIKEGTVTIPYGFIS-----VAANISL--PNYTELYLE 283  
DB 208 IWEVYLSPMWDYVAYKSKENILQGVKIPINBELKHLRTILANNPSRIITPEXFDQMYVT 267  
QY 284 DVEIGTEGTPSTSAHLEWITNITL 310  
DB 268 VWEIGTEFGDPTTTEARFGWTFNSFDI 294

RESULT 12  
AAV17865  
ID AAV17865 standard; protein; 332 AA.  
XX AC AAV17865;  
XX DT 17-AUG-1999 (first entry)  
XX DE Sulfolobus solfataricus endo-beta-1,4-glucanase Cella.  
XX KW Sulfolobus solfataricus; endoglucanase; Cella; CelB;  
XX OS endo-beta-1,4-glucanase.  
XX OS Sulfolobus solfataricus.  
XX FH Key Location/Qualifiers  
XX FT Peptide 1..23  
XX FT /label= signal  
XX FT Protein 24..332  
XX FT /label= Cella  
XX PN DK9900097-A.  
XX PD 12-JAN-1999.  
XX PF 12-JAN-1999; 99DK-000000097.  
XX PR 12-JAN-1999; 99DK-000000097.  
XX PA (NOVO ) NOVO-NORDISK AS.  
XX WPI; 1999-279376/24.  
XX DR N-PSDB; AAX80187.  
XX PT Novel endonuclease - produced by Sulfolobus solfataricus.  
XX PS Claim 1; Page 27-28; 36pp; English.  
XX CC The present sequence represents a Sulfolobus solfataricus endoglucanase,  
XX CC specifically endo-beta-1,4-glucanase, designated Cella. The new endo-beta-  
XX CC 1,4-glucanase is produced by Sulfolobus solfataricus ATCC 35092  
XX SQ Sequence 332 AA;

Query Match 25.8%; Score 439.5; DB 2; Length 332;  
Best Local Similarity 33.4%; Pred. No. 2.3e-32;  
Matches 117; Conservative 57; Mismatches 121; Indels 55; Gaps 17;  
QY 1 MSKKKVVIVSLTLLV-----QAIYFVEKYHTSBDKSTSN-----TSSTP 41  
DB 1 MNKLVIIIVPIIVGVIGGAIYL---HHQSNVKTSSITVTNETTILMSITTNVP 57  
QY 42 PQTLLSTTKVLKIRYDDGWEPCAPIDKDGNGPEFYIEINLWNLNATGFAEMTNLTSL 101  
DB 58 TTVPTTSSIQLIYVTSASSPTFVYLNNTSVFSFYLEVNMNNAKTWNGNYTVVFNPLT 117  
QY 102 GVLHVYQOLDINIVLRDSNVWHGYEPIFYGNKPWNAYATDGPILPISKVSNLT----DP 157  
DB 118 RTLSVSFNLTVQ--NPLQWNTNGYFEIYVGRKPDTSYA--GNI--FPMRIGNMTPEVMVF 171  
QY 158 YLTISYKLEPKNGLPINF--AIESWLTRAWR---TTGINSDEQEVMIWYDGLQAPGS 212  
DB 172 YINLT-KLDPS-----INFDIASDAWIVRQIAFSGTAPGNGDIEIMVWLFSQLQAPG 226  
QY 213 KYKEIVVPIIVNGTPVNAATPEWK-ANI---GWEYVAFRIKTP-----IKEGTVTIPYCAP 264  
DB 227 QVGEVVIPIYINHTLVNATFQWKMKNVPGWCGWBYIAFR---PDGKVTNGYVAYEPNLF 283  
QY 265 ISVAANISLNPYTELXLEDVEIGTEGTPSTSAHLEWITNITLTEL 313  
DB 284 IKALNFASY-NITNYLTDWBEFGTEWGTMTSNGTAIFYSWTISNFYETLL 332

RESULT 13  
AAV17866  
ID AAV17866 standard; protein; 332 AA.  
XX AC AAV17866;  
XX DT 17-AUG-1999 (first entry)  
XX DE Sulfolobus solfataricus endo-beta-1,4-glucanase CelB.  
XX KW Sulfolobus solfataricus; endoglucanase; Cella; CelB;  
XX OS endo-beta-1,4-glucanase.  
XX OS Sulfolobus solfataricus.  
XX FH Key Location/Qualifiers  
XX FT Peptide 1..21  
XX FT /label= signal  
XX FT Protein 22..332  
XX FT /label= CelB  
XX PN DK9900097-A.  
XX PD 12-JAN-1999.  
XX PF 12-JAN-1999; 99DK-000000097.  
XX PR 12-JAN-1999; 99DK-000000097.  
XX PA (NOVO ) NOVO-NORDISK AS.  
XX WPI; 1999-279376/24.  
XX DR N-PSDB; AAX80188.  
XX PT Novel endonuclease - produced by Sulfolobus solfataricus.  
XX PS Claim 1; Page 31-32; 36pp; English.  
XX CC The present sequence represents a Sulfolobus solfataricus endoglucanase,  
XX CC specifically endo-beta-1,4-glucanase, designated CelB. The new endo-beta-  
XX CC 1,4-glucanase is produced by Sulfolobus solfataricus ATCC 35092  
XX SQ Sequence 332 AA;

Query Match 25.3%; Score 430; DB 2; Length 332;  
Best Local Similarity 32.2%; Pred. No. 1.8e-31;  
Matches 111; Conservative 60; Mismatches 124; Indels 50; Gaps 15;  
QY 4 KKFVIVILILLV-----QAIYFVEKHTSEKTSNTSTP-----PQTTL 46  
DB 3 KLIVLEPVIIVIIAIGVMGGIYLHQOQLSVKPVTTTETSTTTTNAITTTVQTVT 62  
QY 47 STTKVLKIRYPDGGEWGPAPIDKDGNGPPEFYIEINLWNLNATGPAEMTYNLTSGVLHY 106  
DB 63 SITSYNQLIIVTSASSEPFPVYNNSTIPSYFLEVNMNKNYNGYTWVFNPLARTLSV 122  
QY 107 VQQLDNIVLRDRSNWVHGYPPIFYGNKPNWANYATDGPILPQSKVNLT----DFYLTIS 162  
DB 123 SFNLTQV---KPLEWTNGYPEIYVGRKPDYTA--GNI--FPMKIGNMTPEWVSFYINLT 176  
QY 163 YKLEPKNGLPIN--ATESMLTAEWR---TTGINSDEQEVMTWYYDGLQAGSKVKEI 217  
DB 177 -KLDPS-----INFDIASDAWIVRQIAFSPGTAAGNGDIEIMVLFSONLQPAGEOVGRV 231  
QY 218 VVPIIVNGTPVNAFFVEVWKANI-----GWYVAVAFRIKTP-----IKEGTVTIPYGAFISVAA 269  
DB 232 VVPIIINHNLVNAIFQVWEMKSPWGGWEYIAFR---PDGKVTNGVVSVEPNLFIKALS 288  
QY 270 NISSLPNYTELYLEDVIEIGTPEGT-PSTTSAHLEWITNITLPL 313  
DB 289 NFTSY-NITNYLYLDWFEFGTEWGTMTSNGTAYFSWTVSFSETLL 332

## RESULT 14

AAW34985  
ID AAW34985 standard; protein; 841 AA.  
AC AAW34985;  
XX  
XX  
DT 27-AUG-2003 (revised)  
DT 21-MAY-1998 (first entry)  
XX  
XX  
DE Archaeobacterial thermostable endoglucanase.  
XX  
XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;  
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;  
KW archaeal bacterium; thermostable enzyme; thermophilic.  
XX  
XX archaeon.  
OS Unidentified.  
XX  
XX WO9744361-A1.  
XX  
XX 27-NOV-1997.  
XX  
XX 22-MAY-1997; 97WO-US008793.  
XX  
XX 22-MAY-1996; 96US-00651572.  
XX  
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.  
XX  
XX Lam DE, Mathur EJ;  
XX  
XX WPI; 1998-018435/02.  
DR N-PSDB; AAT94193.

XX Endoglucanase(s), preferably from archaeal bacterium, AEPII 1a - useful  
XX to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic  
XX bonds in cellulose.

XX Claim 1; Fig 1A; 164pp; English.

XX This protein comprises a thermostable 60.9 kDa endoglucanase of archaeal  
XX bacterial strain AEPII1a, a thermophilic isolate of a marine hydrothermal  
XX vent. The endoglucanase is capable of degrading carboxymethylcellulose  
XX and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It can be  
XX produced from native cells or from recombinant host cells, especially

CC prokaryotic host cells transformed with a plasmid or virus-derived vector  
CC including the endoglucanase DNA (see AAT94193). 23 Other homologous  
CC endoglucanases (see AAW34986-W35008) are also claimed. The endoglucanases  
CC can be used to degrade cellulose for the conversion of plant biomass into  
CC fuels and chemicals, for use in detergents, textiles, animal feed, waste  
CC treatment, and in the fruit juice and brewing industries for the  
CC clarification and extraction of juices. (Updated on 27-AUG-2003 to  
CC correct OS field.)  
XX  
XX

SQ Sequence 841 AA;

Query Match 19.3%; Score 328; DB 2; Length 841;  
Best Local Similarity 67.4%; Pred. No. 2.3e-21;  
Matches 58; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 28 TSEKSTNTSTSTPTQTLLSTTKVLKIRYPDGEWGPAPIDKDGNGPPEFYIEINLWNL 87  
DB 756 TTTQSTTTTTSPTTTTAPQDVTKLRYPDGQWPEAPIDRDKGNGPPEFYIEINPWNIL 815  
QY 88 NATGPAEMTYNLTSGVLHYVQQLDNI 113  
DB 816 SAESYAEMTYNLTSSGVLHYVQQLDSI 841

## RESULT 15

AAW50977  
ID AAW50977 standard; protein; 243 AA.  
XX  
XX AAW50977;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
XX Truncated thermostable cellulase Cell12A (aal9-261).  
DE  
XX Cellulase; Cell12A; thermostable; enzyme; mutant; mutein.  
KW  
XX Rhodothermus marinus.  
OS Synthetic.  
OS  
XX WO200196382-A2.  
XX  
XX 20-DEC-2001.  
XX  
XX 15-JUN-2001; 2001WO-IS000012.  
XX  
XX 15-JUN-2000; 2000US-00594884.  
XX  
XX (PROK-) PROKARIA EHF.  
XX  
XX Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;  
XX  
XX WPI; 2002-226850/28.  
DR N-PSDB; ABA91885.  
XX  
XX Novel thermostable variant cellulase which is truncated such that amino  
XX terminal hydrophobic region and linker group of corresponding full length  
XX enzyme is deleted, has improved catalytic properties and/or stability.  
XX  
XX Claim 15; Page; 51pp; English.

XX The present sequence corresponds to amino acids 19-261 of Rhodothermus  
XX marinus thermostable cellulase Cell12A. This claimed truncated variant of  
XX Cell12A therefore lacks the N-terminal hydrophobic region and 1 of the  
XX amino acids of the linker moiety of the native enzyme (see also  
XX AAW50975). 21 Truncated variants of Cell12A (see AAW50976-996), and  
XX isolated nucleic acids encoding them (see ABA91884-904), are claimed. The  
XX variants have a half-life of at least 3.5 hours at 90 degrees C. They are  
XX more soluble than the native enzyme, and have a specific activity that is  
XX at least 2 times greater than that of the native enzyme. Fusion proteins  
XX comprising a variant thermostable cellulase and a fusion partner are also  
XX claimed, as well as host cells and methods of production. The  
XX thermostable variant cellulases are useful e.g. in the wood and paper  
XX pulp industries, in detergent compositions, in the textile industry, for

CC improving the feed value of animal feedstuffs, and in the food industry.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from the Cell12A sequence given in the Sequence Listing (see  
 CC AAM50975)

XX SQ Sequence 243 AA;

Query Match 9.8%; Score 166.5; DB 5; Length 243;  
 Best Local Similarity 26.6%; Pred. No. 4.8e-07;  
 Matches 66; Conservative 35; Mismatches 108; Indels 39; Gaps 11;  
 QY 59 DGEWPGAPIDKDGNGPEFYIEI-----NLWNLNATCFAMTYNLATSGVLHYV 107  
 Db 1 DWLFPDGDNGKEPEPEPTVELCGRWDRDARDVAGGRYRVINNVWGAETAOCIEVGL---- 56  
 QY 108 QQLDNIVL----RDRSNWVHGYPFIYGNKPNWANYATDGPILPSKVSNLTDFTYLTISY 163  
 Db 57 -ETGNFTITRADHDNGNNVAAYPAYFGCH-WGACTSNG--LPRRVQELSD--VRTSW 109  
 QY 164 KLEPANGLPINFAIESWLTREAWETTGTGINSDEQEVMIWYYD-GLOPAGSKVKEIVVPIL 222  
 Db 110 TLTPITTRWNAAYDIWFSPVTSNGY-SGGAEIMWLNWNGVMPGGRVATVEL--- 165  
 QY 223 VNGTPVNAATFEWVKANIGWEYVAFRIKTIKEGTVTIPYGAFISVAANISSLPNYTELYL 282  
 Db 166 -----AGATWEVWYADWDWNYAVYRTTP-TTSVSELDLKAVIDDAVARGYI--RPEWYL 217  
 QY 283 EDVEIGTE 290  
 Db 218 HAVETGFE 225

Search completed: June 29, 2004, 20:11:34  
 Job time : 63 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 14:39:56 ; Search time 6088 Seconds  
(without alignments)  
6834.644 Million cell updates/sec

Title: US-09-914-543-45  
Perfect score: 960  
Sequence: 1 atgagcagaagaaagtcgt.....atagacctctatttctctaa 960

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hg.\*

3: gb\_in.\*

4: gb\_em.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	960	100.0	960	6	AR205123	Sequence
2	960	100.0	960	6	BD056827	A thermos
3	960	100.0	960	6	BD064594	Glycosida
4	960	100.0	11753	1	AE010200	Pyrococcu
5	958.4	99.8	1134	1	AF181032	Sequence
6	958.4	99.8	1134	6	AX339683	Sequence
7	114.2	11.9	867	6	BD023645	Novel end
8	89.2	9.3	2803	1	TU93354	Thermotoga
9	84.2	8.8	2800	1	TMCELAB	T.maritima
10	84.2	8.8	20553	1	AE001800	Thermotog
11	82	8.5	4547	1	TN286103	T.neapolita
12	53.8	5.6	7218	6	166494	Sequence 14
13	47.4	4.9	2000	6	AX655393	Sequence
14	47.2	4.9	11595	1	AE006803	Sulfolobu
15	46.4	4.8	37037	3	CEP17C11	Caenorhabdi
16	46.2	4.8	128683	2	AC135793	Oryza sat
17	44.2	4.6	13262	1	AE006749	Sulfolobu
18	44	4.6	128683	2	AC135793	Oryza sat
19	43.6	4.5	53561	1	AE000790	Borrelia
20	42.8	4.5	193018	2	BX324226	Danio rer
21	42.4	4.4	1141	6	AX083744	Sequence
22	42.2	4.4	184730	2	AC146676	Papio anu
23	41	4.3	134971	2	AC116367	Oryza sat
24	40.6	4.2	166885	5	BX005129	Zebrafish
25	40.6	4.2	211342	2	BX511164	Danio rer
26	40.2	4.2	144781	9	HSU585114	Human DNA
27	39.6	4.1	129332	9	AL1590028	Human DNA
28	39.6	4.1	149646	2	AL163536	Homo sapi
29	39.6	4.1	160708	2	AC068406	Homo sapi
30	39.6	4.1	163603	2	AC026755	Homo sapi
31	39.6	4.1	197850	2	AC027690	Homo sapi
32	39.6	4.1	201823	9	HSGL20K12	Human DNA
33	39.4	4.1	132202	9	AC025218	Homo sapi
34	39.4	4.1	184772	2	AC114607	Mus muscu
35	39.4	4.1	188627	10	AL844221	Mouse DNA
36	39.4	4.1	194042	2	BX890633	Mus muscu
37	39.4	4.1	265693	2	AC122319	Mus muscu
38	39.4	4.1	274954	2	AC131556	Mus muscu
39	39.2	4.1	239840	2	AC094367	Rattus no
40	39.2	4.1	253095	2	AC130143	Rattus no
41	38.8	4.0	91202	2	AC007609	Homo sapi
42	38.4	4.0	11892	10	AL929538	Mouse DNA
43	38.4	4.0	154864	9	AC143356	Pan trogl
44	38.4	4.0	256680	2	AC098136	Rattus no
45	38.2	4.0	16076	3	AB055862	Procambur

## ALIGNMENTS

RESULT 1	AR205123	Sequence 60 from patent US 6368844.	960 bp	DNA	linear	PAT 20-JUN-2002
LOCUS	AR205123	Sequence 60 from patent US 6368844.				
DEFINITION	AR205123	Sequence 60 from patent US 6368844.				
ACCESSION	AR205123	Sequence 60 from patent US 6368844.				
VERSION	AR205123.1	GI:21502626				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 960)					
AUTHORS	Bylina E.J.					
TITLE	Glycosidase enzymes					
JOURNAL	Patent: US 6368844-A 60 09-APR-2002;					
FEATURES	Location/Qualifiers					

source		1. .960		/organism="unknown"		/mol_type="unassigned DNA"	
ORIGIN		Query Match		Best Local Similarity		Matches	
100.0%; Score 960; DB 6; Length 960;		100.0%; Pred. No. 3.7e-255;		0; Mismatches 0; Indels 0; Gaps 0;		0; Mismatches 0; Indels 0; Gaps 0;	
1		1		1		1	
1		1		1		1	
61		61		61		61	
61		61		61		61	
121		121		121		121	
121		121		121		121	
181		181		181		181	
181		181		181		181	
241		241		241		241	
241		241		241		241	
301		301		301		301	
301		301		301		301	
361		361		361		361	
361		361		361		361	
421		421		421		421	
421		421		421		421	
481		481		481		481	
481		481		481		481	
541		541		541		541	
541		541		541		541	
601		601		601		601	
601		601		601		601	
661		661		661		661	
661		661		661		661	
721		721		721		721	
721		721		721		721	
781		781		781		781	
781		781		781		781	
841		841		841		841	
841		841		841		841	
901		901		901		901	
901		901		901		901	
960		960		960		960	
960		960		960		960	

RESULT 2		BD056827		960 bp		DNA		linear		PAT 27-AUG-2002	
BD056827		A thermostable endo-beta-1,4-glucanase.		BD056827		A thermostable endo-beta-1,4-glucanase.		BD056827		A thermostable endo-beta-1,4-glucanase.	
DEFINITION		BD056827		BD056827		BD056827		BD056827		BD056827	
ACCESSION		BD056827		BD056827		BD056827		BD056827		BD056827	
VERSION		1		1		1		1		1	
KEYWORDS		JP 2001504352-A/1.		JP 2001504352-A/1.		JP 2001504352-A/1.		JP 2001504352-A/1.		JP 2001504352-A/1.	
SOURCE		synthetic construct		synthetic construct		synthetic construct		synthetic construct		synthetic construct	
ORGANISM		artificial sequences.		artificial sequences.		artificial sequences.		artificial sequences.		artificial sequences.	
REFERENCE		1 (bases 1 to 960)		1 (bases 1 to 960)		1 (bases 1 to 960)		1 (bases 1 to 960)		1 (bases 1 to 960)	
AUTHORS		Andersen,L.N., Bjornvad,M.E. and Schulein,M.		Andersen,L.N., Bjornvad,M.E. and Schulein,M.		Andersen,L.N., Bjornvad,M.E. and Schulein,M.		Andersen,L.N., Bjornvad,M.E. and Schulein,M.		Andersen,L.N., Bjornvad,M.E. and Schulein,M.	
TITLE		A thermostable endo-beta-1,4-glucanase		A thermostable endo-beta-1,4-glucanase		A thermostable endo-beta-1,4-glucanase		A thermostable endo-beta-1,4-glucanase		A thermostable endo-beta-1,4-glucanase	
JOURNAL		Patent: JP 2001504352-A 1 03-APR-2001;		Patent: JP 2001504352-A 1 03-APR-2001;		Patent: JP 2001504352-A 1 03-APR-2001;		Patent: JP 2001504352-A 1 03-APR-2001;		Patent: JP 2001504352-A 1 03-APR-2001;	
COMMENT		NOVO NORDISK AS		NOVO NORDISK AS		NOVO NORDISK AS		NOVO NORDISK AS		NOVO NORDISK AS	
PN		JP 2001504352-A/1		JP 2001504352-A/1		JP 2001504352-A/1		JP 2001504352-A/1		JP 2001504352-A/1	
PD		03-APR-2001		03-APR-2001		03-APR-2001		03-APR-2001		03-APR-2001	
PF		30-JAN-1998 JP 1998532468		30-JAN-1998 JP 1998532468		30-JAN-1998 JP 1998532468		30-JAN-1998 JP 1998532468		30-JAN-1998 JP 1998532468	
PR		31-JAN-1997 DK 0114/97,11-JUL-1997 DK 0853/97 PI		31-JAN-1997 DK 0114/97,11-JUL-1997 DK 0853/97 PI		31-JAN-1997 DK 0114/97,11-JUL-1997 DK 0853/97 PI		31-JAN-1997 DK 0114/97,11-JUL-1997 DK 0853/97 PI		31-JAN-1997 DK 0114/97,11-JUL-1997 DK 0853/97 PI	
LENE NONBOE ANDERSEN,MADS ESKELUND BJORNVAD,MARTIN SCHULEIN PC		LENE NONBOE ANDERSEN,MADS ESKELUND BJORNVAD,MARTIN SCHULEIN PC		LENE NONBOE ANDERSEN,MADS ESKELUND BJORNVAD,MARTIN SCHULEIN PC		LENE NONBOE ANDERSEN,MADS ESKELUND BJORNVAD,MARTIN SCHULEIN PC		LENE NONBOE ANDERSEN,MADS ESKELUND BJORNVAD,MARTIN SCHULEIN PC		LENE NONBOE ANDERSEN,MADS ESKELUND BJORNVAD,MARTIN SCHULEIN PC	
C12N9/42		C12N9/42		C12N9/42		C12N9/42		C12N9/42		C12N9/42	
CC		Strandedness: Single;		Strandedness: Single;		Strandedness: Single;		Strandedness: Single;		Strandedness: Single;	
CC		Topology: Linear;		Topology: Linear;		Topology: Linear;		Topology: Linear;		Topology: Linear;	
FH		Key		Key		Key		Key		Key	
Location/Qualifiers.		Location/Qualifiers.		Location/Qualifiers.		Location/Qualifiers.		Location/Qualifiers.		Location/Qualifiers.	
source		1. .960		1. .960		1. .960		1. .960		1. .960	
/organism="synthetic construct"		/organism="synthetic construct"		/organism="synthetic construct"		/organism="synthetic construct"		/organism="synthetic construct"		/organism="synthetic construct"	
/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"	
/db_xref="taxon:32630"		/db_xref="taxon:32630"		/db_xref="taxon:32630"		/db_xref="taxon:32630"		/db_xref="taxon:32630"		/db_xref="taxon:32630"	
ORIGIN		Query Match		Best Local Similarity		Matches		Query Match		Best Local Similarity	
100.0%; Score 960; DB 6; Length 960;		100.0%; Pred. No. 3.7e-255;		0; Mismatches 0; Indels 0; Gaps 0;		0; Mismatches 0; Indels 0; Gaps 0;		100.0%; Score 960; DB 6; Length 960;		100.0%; Pred. No. 3.7e-255;	
1		1		1		1		1		1	
1		1		1		1		1		1	
61		61		61		61		61		61	
61		61		61		61		61		61	
121		121		121		121		121		121	
121		121		121		121		121		121	
181		181		181		181		181		181	
181		181		181		181		181		181	
241		241		241		241		241		241	
241		241		241		241		241		241	
301		301		301		301		301		301	
301		301		301		301		301		301	
361		361		361		361		361		361	
361		361		361		361		361		361	
421		421		421		421		421		421	
421		421		421		421		421		421	
481		481		481		481		481		481	
481		481		481		481		481		481	
541		541		541		541		541		541	
541		541		541		541		541		541	
601		601		601		601		601		601	
601		601		601		601		601		601	
661		661		661		661		661		661	
661		661		661		661		661		661	
721		721		721		721		721		721	
721		721		721		721		721		721	
781		781		781		781		781		781	
781		781		781		781		781		781	
841		841		841		841		841		841	
841		841		841		841		841		841	
901		901		901		901		901		901	
901		901		901		901		901		901	
960		960		960		960		960		960	
960		960		960		960		960		960	



Db 541 TTAACGAGAGAGCTTGGAGAACACACAGGAATTAACAGCGATGACGAAGTAATGATA 600  
Qy 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660  
Db 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660  
Qy 661 ATAATAGTAAACCGAAACACGAGTAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720  
Db 661 ATAATAGTAAACCGAAACACGAGTAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720  
Qy 721 TGGAGTATGCTGCATTTAGATTAAGACACCCCAATCAAGAGGGAACAGTGACAATCCA 780  
Db 721 TGGAGTATGCTGCATTTAGATTAAGACACCCCAATCAAGAGGGAACAGTGACAATCCA 780  
Qy 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTAACAGAACTT 840  
Db 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTAACAGAACTT 840  
Qy 841 TACTTAGAGACGTTGGAGATTGGAACTGGATTTGGAAAGCCCAAGCACTACTCGGCCAC 900  
Db 841 TACTTAGAGACGTTGGAGATTGGAACTGGATTTGGAAAGCCCAAGCACTACTCGGCCAC 900  
Qy 901 CTAGAGTGGTGGATACAAACATTAACACTAACTACTCTCTAGATAGACCTCTTATTCTTAA 960  
Db 901 CTAGAGTGGTGGATACAAACATTAACACTAACTACTCTCTAGATAGACCTCTTATTCTTAA 960

RESULT 3  
BD064594  
LOCUS BD064594 960 bp DNA linear PAT 27-AUG-2002  
DEFINITION Glycosidase enzymes.  
ACCESSION BD064594.1 GI:22610197  
VERSION JP 2001505440-A/46.  
KEYWORDS Vaccinia virus  
SOURCE ORGANISM  
Vaccinia virus  
Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
Orthopoxvirus.  
REFERENCE 1 (bases 1 to 960)  
BYLINE E.J., Swanson, R.V., Mathur, E.J. and Lam, D.E.  
TITLE Glycosidase enzymes  
JOURNAL Patent: JP 2001505440-A 46 24-APR-2001;  
DIVERSA CORP  
COMMENT PN JP 2001505440-A/46  
PD 24-APR-2001  
PF 08-DEC-1997 JP 1998525887  
PR 06-DEC-1996 US 60/056916  
PI EDWARD J BYLINE, RONALD V SWANSON, ERIC J MATHUR, DAVID E LAM PC  
C07H21/04, C12N1/20, C12N1/14, C12N5/00, C12N9/38, C12N9/42, C08B30/ PC  
04  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers.

FEATURES  
source  
1..960  
/organism="Vaccinia virus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10245"

ORIGIN  
Query Match 100.0%; Score 960; DB 6; Length 960;  
Best Local Similarity 100.0%; Pred. No. 3.7e-255;  
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGCAGAAAAGTTGTCATCGTATCTTATCTTATCAATCTTTTAGTACAGGCAATA 60  
Db 1 ATGAGCAGAAAAGTTGTCATCGTATCTTATCTTATCAATCTTTTAGTACAGGCAATA 60  
Qy 61 TATTTTGTAGAAAAGTATCATCTCTGAGGACAGTCAACTTCAAAATACCTCATCTACA 120  
Db 61 TATTTTGTAGAAAAGTATCATCTCTGAGGACAGTCAACTTCAAAATACCTCATCTACA 120

Qy 121 CCACCCCAAAACACACTTTCCACTACCAAGGTTCTCAAGATTAGTATACCTTGATGACGGT 180  
Db 121 CCACCCCAAAACACACTTTCCACTACCAAGGTTCTCAAGATTAGTATACCTTGATGACGGT 180  
Qy 181 GAGTGGCCAGGAGCTTCCTATTGATAAGGATGGTGTGTAAGGAAACCCAGAAATCTTACATTGAA 240  
Db 181 GAGTGGCCAGGAGCTTCCTATTGATAAGGATGGTGTGTAAGGAAACCCAGAAATCTTACATTGAA 240  
Qy 241 ATAAACCTTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACAAATTTAAC 300  
Db 241 ATAAACCTTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACAAATTTAAC 300  
Qy 301 ACGCGGCTCTTCACTAGTCCCAACATTTGCAACACTTGTCTTGAGGATAGAAAGTAAAT 360  
Db 301 ACGCGGCTCTTCACTAGTCCCAACATTTGCAACACTTGTCTTGAGGATAGAAAGTAAAT 360  
Qy 361 TGGGTGCTATGATACCCGCAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420  
Db 361 TGGGTGCTATGATACCCGCAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420  
Qy 421 ACTGATGGCCCAATACCATTAACCAAGTAAAGTTTCAAACTTAACAGACTTCTATCTAACA 480  
Db 421 ACTGATGGCCCAATACCATTAACCAAGTAAAGTTTCAAACTTAACAGACTTCTATCTAACA 480  
Qy 481 ATCTCTATAAACTTTGAGCCCAAGAACCGCTTGCCTTAACTTCGCAATAGAAATCTCTGG 540  
Db 481 ATCTCTATAAACTTTGAGCCCAAGAACCGCTTGCCTTAACTTCGCAATAGAAATCTCTGG 540  
Qy 541 TTAACGAGAGAGCTTTGGAGAACAAACAGAAATTAACAGCGATGACGAAGTAATGATA 600  
Db 541 TTAACGAGAGAGCTTTGGAGAACAAACAGAAATTAACAGCGATGACGAAGTAATGATA 600  
Qy 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660  
Db 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660  
Qy 661 ATAATAGTAAACCGAAACACAGTAATGCTACATTTGAAGTATGGAAGCAAACTTGGT 720  
Db 661 ATAATAGTAAACCGAAACACAGTAATGCTACATTTGAAGTATGGAAGCAAACTTGGT 720  
Qy 721 TGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGGAACAGTGACAATCCA 780  
Db 721 TGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGGAACAGTGACAATCCA 780  
Qy 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTAACAGAACTT 840  
Db 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTAACAGAACTT 840  
Qy 841 TACTTAGAGACGTTGGAGATTGGAACTGGATTTGGAAAGCCCAAGCACTACTCGGCCAC 900  
Db 841 TACTTAGAGACGTTGGAGATTGGAACTGGATTTGGAAAGCCCAAGCACTACTCGGCCAC 900  
Qy 901 CTAGAGTGGTGGATCACAACATTAACACTAACTACTCTCTAGATAGACCTCTTATTCTTAA 960  
Db 901 CTAGAGTGGTGGATCACAACATTAACACTAACTACTCTCTAGATAGACCTCTTATTCTTAA 960

RESULT 4  
AE010200/c  
LOCUS AE010200 11753 bp DNA linear BCT 25-FEB-2002  
DEFINITION Pyrococcus furiosus DSM 3638, section 75 of 173 of the complete genome.  
ACCESSION AE010200 AE009950  
VERSION AE010200.1 GI:18892881  
KEYWORDS Pyrococcus furiosus DSM 3638  
SOURCE Pyrococcus furiosus DSM 3638  
ORGANISM Pyrococcus furiosus DSM 3638  
REFERENCE 1 (bases 1 to 11753)  
AUTHORS Maeder, D.L., Weiss, R.B., Dunn, D.M., Cherry, J.L., Gonzalez, J.M., DiRuggiero, J. and Robb, F.T.  
TITLE Divergence of the hyperthermophilic archaea Pyrococcus furiosus and

P. horikoshii inferred from complete genomic sequences  
Genetics 152 (4), 1299-1305 (1999)

JOURNAL MEDLINE 99359404  
PUBMED 10430560  
REFERENCE 2 (bases 1 to 11753)  
AUTHORS Robb,F.T., Maeder,D.L., Brown,J.R., DiRuggiero,J., Stump,M.D.,  
Yeh,R.K., Weiss,R.B. and Dunn,D.M.  
TITLE Genomic sequence of hyperthermophile, *Pyrococcus furiosus*:  
implications for physiology and enzymology  
Meth. Enzymol. 330, 134-157 (2001)

JOURNAL MEDLINE 21079003  
PUBMED 11210495  
REFERENCE 3 (bases 1 to 11753)  
AUTHORS Weiss,R.B., Dunn,D.M., Robb,F.T. and Brown,J.R.  
TITLE The complete sequence of the *Pyrococcus furiosus* genome  
Unpublished

JOURNAL MEDLINE 21079003  
PUBMED 11210495  
REFERENCE 4 (bases 1 to 11753)  
AUTHORS Weiss,R.B.  
TITLE Direct Submission  
JOURNAL Submitted (12-FEB-2002) Human Genetics, University of Utah, 20  
South 2030 East, Salt Lake City, UT 84112, USA  
FEATURES Location/Qualifiers  
1..11753  
/organism="Pyrococcus furiosus DSM 3638"  
/mol\_type="genomic DNA"  
/strain="DSM 3638"  
/db\_xref="taxon:186497"  
166..1242  
/gene="PF0849"  
166..1242  
/gene="PF0849"  
166..1242  
/note="Function Code: 16.1 Conserved Hypothetical"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAL80973.1"  
/db\_xref="GI:18892882"  
/translation="MYELTDFDKLRKITYKLGVDREDLLVIPSSKAGPCGNCL  
FLYLLQPPENIVYVARDHLNDPTLEERIRYARKHYDLWIRVDTSGNVKFDENIK  
SCYAGLDDEIOISVHTTKQVRIKLMNRHAGKIDLLPLVAKHFRTIADIILTPGFN  
VDIGIEIDDSMGVEHVRPLPVGVTKYRFEIRPLTKELSVMKVEALEKDELGI  
KVPVPIPLFALLGFTTGLPEPNIPEPFTYITGELAYEMKRLPFIKIVAVKNEF  
FCGNIGAGLTGRLVREVERLPEVDFGLILLPELMFYGDMLDGNRRQDLFSKILI  
EKGIVTALPEPFIKIVKISL"  
1235..1289  
/note="small nucleolar RNA, posttranscriptional  
modification of rRNA"  
1265..2380  
/gene="PF0850"  
1265..2380  
/gene="PF0850"  
/note="Function Code: 16.1 Conserved Hypothetical"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAL80974.1"  
/db\_xref="GI:18892883"  
/translation="MIRGLAEGRMKETSIVYVGEPLPKGCQYCVRGKULVFTG  
VCPNCPYCLSPWRKDVAYANERPIRKIEDIIEAKIQEARGAGVTGGDPLARLN  
TVEYIRLKEEFKGFHLYTGILADKRALMLYDAGLDEIRFHPDLPSPSKFLN  
REIENLNKAPFDWDVGVEVPVPGFGRDKWFAEILDKFGKFLNILEYSEITLS  
NILNRGFKLAGENFAIKGSLGLEILEWKENKTSLYHLCTAKLKDAVOLNRLLR  
MAKNVAPYWEITDGLRGLAEVDELELYKFLVEEAEPVKENLYINWEKRIEMP  
VEVAEELAELEGVDFYVEEPTWDRIEVERIPLN"  
2459..2860  
/gene="PF0851"  
2459..2860  
/gene="PF0851"  
/note="Function Code: 10 Regulatory Functions: REGULATORY  
FUNCTIONS"  
/codon\_start=1  
/transl\_table=1  
/product="iron-dependent repressor"

gene  
CDS

snRNA

gene  
CDS

gene  
CDS

gene  
CDS

protein\_id="AAL80975.1"  
/db\_xref="GI:18892884"  
/translation="MYIQKNGVIRVDKIAMMRVKPPTVVEALKKLRDKGVKVEE  
HEHILLTEGLEVAKTKYKQLLTFFINILGIPPEIAERDQCFHYSEVTVHRI  
REFISYIQECPYALKQFLKVKREKQAVAK"  
complement(2842..4170)  
/gene="PF0852"  
complement(2842..4170)  
/gene="PF0852"  
/note="Function Code: 16.1 Conserved Hypothetical"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAL80976.1"  
/db\_xref="GI:18892885"  
/translation="MGWIERFEPEKCYGTDLKTVELAGITTFMTMAIIFVNPILSN  
AMGPEAFSLVAATSLAAGLTILMGIVYAKPFALPGMGLNAFAFTVAPKYGKVA  
LAAVEGIIIFIIISITKVSAAIHAIPLSQYAGAGIGLFTFGLNNVGLLTALT  
GDNGLLIKTSLNASLAKPEGLFLGLFTAILLSIKGALLISILATSIIIGMITG  
VAPRDEGVSMPTLSYTFQLDLGNAGALGAFVAFPMVDFTLTGTVGLSAKAG  
ELTKEGKIPDAEKVLLTAIGTTCGAILGTSTVTYLSAAGIBEGRTGFTAVITGL  
LELIIGLSIAPVAKAIAPAPATAPALVINGYIMISAKFVDESDTEALPAFLVLTIP  
FYISADGIGVGFISITLKVFTGRKWEVHPMLTALAVVFIGYFAYLATA"  
complement(4316..5113)  
/gene="PF0853"  
complement(4316..5113)  
/gene="PF0853"  
/note="Function Code: 9.5 Purines, Pyrimidines,  
Nucleosides and Nucleotides: Salvage of nucleosides and  
nucleotides"  
/codon\_start=1  
/transl\_table=1  
/product="5'-methythioadenosine phosphorylase II"  
/protein\_id="AAL80977.1"  
/db\_xref="GI:18892886"  
/translation="MPRIATVGGGVYDFPAENKREETVKTGYGEVKITVGVGDEEV  
AFLRKHGHSIPHKINRYANIVAHVYELGVERIIAGSAVSMNPKPGDFVILDQI  
IDFTVSRRTFYDGEESPHERRFVAVDTFYCPYKRAKALITARNLGLDHPRGTY  
VCTGEPRTAAETRAYRILGCVGVMTQCPPEILARELCMECYATVAIVNYAAGMSG  
KKLTSEVSEVMKQKSDIVKLIILAAIPLIKERRCGCKALKGATG"  
5176..5253  
/product="tRNA-Arg"  
/note="tRNA Arg anticodon CCG"  
complement(5400..6359)  
/gene="PF0854"  
complement(5400..6359)  
/gene="PF0854"  
/note="Function Code: 5.1 Central Intermediary Metabolism:  
degradation of polysaccharides"  
/codon\_start=1  
/transl\_table=1  
/product="endo-1,4-beta-glucanase b"  
/protein\_id="AAL80978.1"  
/db\_xref="GI:18892887"  
/translation="MSKKKFIVSITLITLLVQAIYFVEKYHTSDKSTNTSTPPQT  
TLSTTKVLKIRYDDGWEFGAPIDKDGNGPEFYIEINLWILNATGPAEYMYLTSG  
VLHYQOLDNIVLRDSRWVHGYEIFYGNKPNANVATDGPILPSKVSNLATDYLT  
ISYKPEKNGDIPNFAIESWLTREAWRTGINSDEQEWMIWYIDGLQPGASKVEIV  
VPIVNGPVNATVEKNAIGWEYFAPRTPIKEGTVTPIYGFAGISVAANSISLEN  
YTEIYLEVEITGTEGFTSTTSASHLEWITNITILTPDLRLIS"  
complement(6356..7246)  
/gene="PF0855"  
complement(6356..7246)  
/gene="PF0855"  
/note="Function Code: 16.1 Conserved Hypothetical"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAL80979.1"  
/db\_xref="GI:18892888"  
/translation="WAVRSFGAIDFLVKYKLLLLTIIPALJSGVASPFVEEYEPET  
QRFDIIERYGAISPENWKLPIYPIVLLVITLSIGFEGILIPAIKAKGKDKFSL  
KEIAIYURNAGWGLLLTIISLLXLLALVIGGSIPTIILLVDFIVDFLEPLVIVEL

LFVWILILFLSLVSGILVFPFVDRSLRRLRVIKYGRYKSMIGFALLGLSS  
 FALLMAGIVENSVMFSKNTVARLLASLLTPVLIALLSLFVIAETIYENVRADYE  
 ERALFYTLVSNLTNLEVDQ"  
 7996. .8072  
 /product="trna-Ala"  
 /notes="trna Ala anticodon CGC"  
 8097. .9281  
 /gene="PF0856"  
 8097. .9281  
 /gene="PF0856"  
 /notes="Function Code: 16.1 Conserved Hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAL80980.1"  
 /db\_xref="GI:11892889"  
 /translation="MDEGRTVLRKPEHLEHCLKRVNVEAHATNGPEDVHVMSPLEI  
 DKDEIDLSVFLGRKEDYPMITGTTGGTNGKGEVAMKINRTLAQAAELNIFLGVGSQ  
 RAMIEXPTWESYVDRDVAENFLVNLGAPQGRNAKRYGVKVEVLYAIEKIDADIG  
 AIHNPQLQESVQEGDPTFSVLEALAEITSSIDYPIAKETGAGKSVKVAIKLESIG  
 VSADISGVGTGWSGVYRAKDELKRLALFWDWGIKTAISLAVERFSTNLPILIA  
 SGMRDGILTAKALAGASLVGIALPVLPAKGDVGVKIVKIGVBEIKNAMFLVG  
 ARNVEELRKVPIVFTGFVREWLQRIDLOEFVKRAKLRFEEFWNVGVFYPYTSSTH  
 LF"  
 complement(9442. .11433)  
 /gene="PF0857"  
 complement(9442. .11433)  
 /gene="PF0857"  
 /notes="Function Code: 14.5 Transport and Binding Proteins:  
 Cations"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ferrous iron transport protein b"  
 /protein\_id="AAL80981.1"  
 /db\_xref="GI:11892890"  
 /translation="MLKTVLVGNPNVKKTIIFNALTGLRQHVGNVPGVTVEKKEGIM  
 EYREKFLVDLPGIYSLTAHSIDELIARNFILDGNADVIDVDSTCLWRNLFLTE  
 LFEVKNILVNLKFLDKKAKIDIKMRKELGVPIPTNAKKGEGVEELKRMIA  
 LMAEKNVITNPIIPRYDEIERIKHISELLRGTPLAERKPIRWALKLQLQRESEVIR  
 LVLKYLQEQKMEILKHISELEBKYPRLDIVIASQKFEQLLRKFRVHBEVKEN  
 100.0%; Score 960; DB 1; Length 11753;  
 Query Match  
 Best Local Similarity 100.0%; Pred. No. 3.9e-255;  
 Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACCAAGAAAAGTTCGTATCGTATCTATCTTAACTCTTTTAGTACAGGCAATA 60  
 DB 6359 ATGACCAAGAAAAGTTCGTATCGTATCTATCTTAACTCTTTTAGTACAGGCAATA 6300  
 QY 61 TATTTCTGAGAAAGTATACCTCTGAGGACAACTTCAAAATACCTCATCTACA 120  
 DB 6299 TATTTCTGAGAAAGTATACCTCTGAGGACAACTTCAAAATACCTCATCTACA 6240  
 QY 121 CCACCCCAACAACTTCCACTACCAAGTCTTCAAGATTAGATACCTGTGACGGT 180  
 DB 6239 CCACCCCAACAACTTCCACTACCAAGTCTTCAAGATTAGATACCTGTGACGGT 6180  
 QY 181 GAGTGGCCAGAGTCTCTATTGATAAGGATGGATGGACCCAGCAAGTCTTACATTGAA 240  
 DB 6179 GAGTGGCCAGAGTCTCTATTGATAAGGATGGATGGACCCAGCAAGTCTTACATTGAA 6120  
 QY 241 ATAAACCTATGGAACATCTTAAAGTCTGAGTTCGTGAGATGAGTCAATTTAAAC 300  
 DB 6119 ATAAACCTATGGAACATCTTAAAGTCTGAGTTCGTGAGATGAGTCAATTTAAAC 6060  
 QY 301 AGCGGCGTCTTCACTAGCTCCACAACTTGACAACTTGTCTTGAGGGATAGAGTAAT 360  
 DB 6059 AGCGGCGTCTTCACTAGCTCCACAACTTGACAACTTGTCTTGAGGGATAGAGTAAT 6000  
 QY 361 TGGGTGATGATACCCCGAATATCTTATGGAACCAAGCCATGGAATGCAAACTACGCA 420  
 DB 5999 TGGGTGATGATACCCCGAATATCTTATGGAACCAAGCCATGGAATGCAAACTACGCA 5940

QY 421 ACTGATGCCCAATACCACTTACCCAGTAAAGTTTCAAACTAACAGAGTCTTATCTAACA 480  
 DB 5939 ACTGATGCCCAATACCACTTACCCAGTAAAGTTTCAAACTAACAGAGTCTTATCTAACA 5880  
 QY 481 AFTCTCTATAAATTGAGCCCAAGAACCGGCTGCCAATTAATCTTGGCAATAGATCTCTGG 540  
 DB 5879 AFTCTCTATAAATTGAGCCCAAGAACCGGCTGCCAATTAATCTTGGCAATAGATCTCTGG 5820  
 QY 541 TTAACGAGAGAGCTTGGAGAACCAACAGCAATTAACACGCGATGAGCAAGAAGTAAATGATA 600  
 DB 5819 TTAACGAGAGAGCTTGGAGAACCAACAGCAATTAACACGCGATGAGCAAGAAGTAAATGATA 5760  
 QY 601 TGGATTTACTATGACGATTAACACCGGCTGCCAATTAATCTTGGCAATAGATCTTGTAGTCCA 660  
 DB 5759 TGGATTTACTATGACGATTAACACCGGCTGCCAATTAATCTTGGCAATAGATCTTGTAGTCCA 5700  
 QY 661 ATAATAGTTTAACGGAACACCACTTAATCTTGAAGTATGGAAGGCAAAACATTTGGT 720  
 DB 5699 ATAATAGTTTAACGGAACACCACTTAATCTTGAAGTATGGAAGGCAAAACATTTGGT 5640  
 QY 721 TGGGAGTATGTTGCACTTTAGATAAAGACCCCAATCAAAGAGGGAACAGTGAATTTCCA 780  
 DB 5639 TGGGAGTATGTTGCACTTTAGATAAAGACCCCAATCAAAGAGGGAACAGTGAATTTCCA 5580  
 QY 781 TACGAGAGCTTTATTAAGTGTTCGACCAACATTTCAAGCTTACCAAAATACAGAACTT 840  
 DB 5579 TACGAGAGCTTTATTAAGTGTTCGACCAACATTTCAAGCTTACCAAAATACAGAACTT 5520  
 QY 841 TACTTTAGAGGAGCTGAGATTTGGAATGAGTTTGGACCCCAAGCACTACCTCCGCCAC 900  
 DB 5519 TACTTTAGAGGAGCTGAGATTTGGAATGAGTTTGGACCCCAAGCACTACCTCCGCCAC 5460  
 QY 901 CTAGAGTGTGATCACAACATTAACACTTCTCTTAGATAGACCTTATTTTCTTAA 960  
 DB 5459 CTAGAGTGTGATCACAACATTAACACTTCTCTTAGATAGACCTTATTTTCTTAA 5400  
 RESULT 5  
 AF181032 1134 bp DNA linear BCT 05-NOV-1999  
 LOCUS Pyrococcus furiosus endoglucanase A precursor (egia) gene, complete cds.  
 AF181032  
 AF181032.1 GI:5870828  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pyrococcus furiosus  
 Pyrococcus furiosus  
 Archaea; Euryarchaeota; Thermococci; Thermococcales;  
 Thermococcaceae; Pyrococcus.  
 REFERENCE  
 1 (bases 1 to 1134)  
 BYLINA, E.J., SWANSON, R.V., MATHUR, E.J., LAM, D.E., and CALLAN, W.  
 TITLES  
 Glycosidases  
 Patent: USA (08/949,026) 10-OCT-1997; Abdulla  
 DIVERSA CORPORATION; 10665 Sorrento Valley Road; San Diego, CA;  
 USA;  
 REMARK  
 Patent pending  
 REFERENCE  
 2 (bases 1 to 1134)  
 BAUER, M.W., DRISKILL, L.E., CALLEN, W., SNEAD, M.A., MATHUR, E.J., and  
 KELLY, R.M.  
 TITLES  
 An endoglucanase, EglA, from the hyperthermophilic archaeon  
 Pyrococcus furiosus hydrolyzes beta-1,4 bonds in mixed-linkage  
 (1-->3), (1-->4)-beta-D-glucans and cellulose  
 J. Bacteriol. 181 (1), 284-290 (1999)  
 99084968  
 MEDLINE  
 PUBMED  
 9864341  
 REFERENCE  
 3 (bases 1 to 1134)  
 BAUER, M.W., DRISKILL, L.E., and KELLY, R.M.  
 TITLES  
 Direct Submission  
 Submitted (25-AUG-1999) Chemical Engineering, North Carolina State  
 University, P. O. Box 7905, Raleigh, NC 27695, USA  
 JOURNAL  
 REFERENCE  
 4 (bases 1 to 1134)  
 CALLEN, W., SNEAD, M.A., and MATHUR, E.J.  
 TITLES  
 Direct Submission



DEFINITION	Novel endoglucanase.
ACCESSION	B0023645
VERSION	BD023645.1 GI:22564868
KEYWORDS	JP 2001507221-A/1.
SOURCE	unidentified
ORGANISM	unclassified
REFERENCE	1 (bases 1 to 867)
AUTHORS	Schilein,M., Bjornvad,M.E. and Norrevang,I.A.
TITLE	Novel endoglucanase
JOURNAL	Patent: JP 2001507221-A 1 05-JUN-2001;
COMMENT	NOVO NORDISK AS PN JP 2001507221-A/1 PD 05-JUN-2001 PF 19-DEC-1987 JP 1998528260 PR 20-DEC-1996 DK 1483/96 PI MARTIN SCHILEIN,MADS ESKELUND BJORNVAD,I BEN ANGELICA NORREVRANG PC C12N15/09,A23K1/165,C12N1/15,C12N1/19,C12N1/21,C12N9/42 PC C12P7/06,C12P19/00, PC C12S3/00,D06M16/00/A23P5/24,(C12N15/09,C12R1/145),(C12N1/15, C12R1/685), PC C12N1/15,C12R1/69),(C12N1/15,C12R1/77),(C12N1/15,C12R1/885), PC C12N1/19,C12R1/865),(C12N1/21,C12R1/125),(C12N1/21,C12R1/07), PC C12N1/21,C12R1/101,C12N15/00,(C12N15/00,C12R1/145) CC Strandedness: Single; CC Topology: Linear; FH Key Location/Qualifiers. source 1..867 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"
ORIGIN	
Query Match	11.9%; Score 114.2; DB 6; Length 867;
Best Local Similarity	51.3%; Pred. No. 1.6e-20;
Matches 291; Conservative	0; Mismatches 273; Indels 3; Gaps 1;
Qy	230 TCTACATTGAAATAAACCTATGGAACATCTTTAATGCTACTTGGATTGCTGAGATGACGT 289
Db	137 TCACCTTAGAACTCAACTTTTGGAAATATTGCAAACTATGAAGGAAATACATGATGGATGCAT 196
Qy	290 ACAATTTAACCGGGCGTCTTCTACCTACCTCCACCAACTTGCACACATTTGCTCTTGAGG 349
Db	197 TTTATAAGGAAGAAGATACCTGTTGAATATTATGCCGACATAAAAAACATAGTACTTAAG 256
Qy	350 ATAGAAGTAATTGGGTGCATGGATACCCCGAAATATTCTATGGAACCAAGCCATGGAATG 409
Db	257 ATAAAAATTCATGGGTACATGGATATCTCTGAAGTCTACTATGGTACAAACCATGGGCTG 316
Qy	410 CAAACTAGCGAATGATGGCCCAATACCATTAACCCAGTAAAGTTTCAACCTAACAGACT 469
Db	317 GCCATGGGAATTCANTGAGAAATTAGCTCTTCTCTAAAAAGGTATCAGAAATTCACGACG 376
Qy	470 TCTATCTAACAACTCTCTATAAACTTCAGCGCCCAAGAACGGCGCTGCCAATTAACCTTCGCA 529
Db	377 TTCTCTCAATCTAAATAATACAAATATGTTACGAGAGAAATCTTCCATATAATTTTGCTA 436
Qy	530 TAGAATCCTGGTTAAACGAGAGAGCTTTGGAGAACACACAGGAAATTAACAGCGATGACGAC 589
Db	437 TGGAAACATGGATAAACAAAAGAACCCCTATCAGAAAACCG---TTACTTCAGGGGATATAG 493
Qy	590 AAGTAATGATATGATTTACTATGACGGATTACAAACGGCTGGCTCCAAAGTTAAAGGAGA 649
Db	494 AGATGATGGTATGGCTATATGCTATATAGACTTTCTCTCGAGCGCAAGGTAGGAGAG 553
Qy	650 TTGTAGTCCCAATAATAGTTAAACGGAAACACAGTAAATGCTACATTTGAAAGTATGGAAG 709
Db	554 TAAAAATACCTATCATCCTTAAACGGTAAATCAAAAAAGACATTTCTCTGGAAGTATATCTTT 613
Qy	710 CAAACATTTGGTGGAGATATGTTGCATTTAGATATAAGACCCCAATCAAGAGGGACAG 769
Db	614 CCCCTATGAGCTGGGACTAGTGCCCTATTAATCAAAAGAAAATATTTCTTCAAGACAGG 673

```

QY 770 TGACAAATTCATACGAGCATTTATA 796
Db 674 TAAATAACCAATAAATGAATTTTGA 700

RESULT 8
TN093354 2803 bp DNA linear BCT 11-DEC-1998
LOCUS Thermotoga neapolitana endo-1,4-beta-glucanase A (celA) and
DEFINITION endo-1,4-beta-glucanase B (celB) genes, complete cds.
ACCESSION U93354
VERSION U93354.1 GI:1934672
KEYWORDS Thermotoga neapolitana
SOURCE Thermotoga neapolitana
ORGANISM Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
REFERENCE 1 (bases 1 to 2803)
AUTHORS Bok, J.D., Vernool, D.A. and Eveleigh, D.E.
TITLE Purification, characterization, and molecular analysis of
thermostable cellulases CelA and CelB from Thermotoga neapolitana
JOURNAL Appl. Environ. Microbiol. 64 (12), 4774-4781 (1998)
MEDLINE 98054912
PUBMED 9835561
REFERENCE 2 (bases 1 to 2803)
AUTHORS Bok, J.D. and Eveleigh, D.E.
DIRECT SUBMISSION
TITLE Submitted (13-MAR-1997) Biochemistry and Microbiology, Cook
JOURNAL College, Rutgers University, Lipman Drive, New Brunswick, NJ 08903,
USA

FEATURES
source
1..2803
location/Qualifiers
1..2803
/organism="Thermotoga neapolitana"
/mol_type="genomic DNA"
/db_xref="taxon:2337"
1..774
/gene="celA"
1..774
/gene="celA"
/EC_number="3.2.1.4"
/function="hydrolysis of carboxymethyl cellulose, beta
glucan"
/notes="TNCEL A"
/codon_start=1
/evidence-experimental
/transl_table=11
/product="endo-1,4-beta-glucanase A"
/protein_id="AAC95059.1"
/db_xref="GI:1934673"
/translation="MVELTAPGTADFRNDMPISMEINLVNRYGTVMVRPDGRL
TFNGDVEDLAREPRLGYEFYGYKWRHAAAGTKLPLVSVSVSTVLSFE
IDHPSLPLFAMFTWIRKQYQVQASIGDVEIMVWFYFNLTPGKTKGYEYVSEL
NGEHRGRIWELHAEKNNDYLAPLKNPVRKRVRFNVKDFLDVAGEYLSRSTRVKDF
DDLVTWEIGTEFGSPETKSAREGTFNFSIDMEVKG"
759..1595
/gene="celB"
759..1595
/gene="celB"
771..1595
/gene="celB"
/EC_number="3.2.1.4"
/function="hydrolysis of carboxymethyl cellulose, beta
glucan"
/notes="TNCEL B"
/codon_start=1
/evidence-experimental
/transl_table=11
/product="endo-1,4-beta-glucanase B"
/protein_id="AAC95060.1"
/db_xref="GI:1934674"
/translation="MRVLVSVFLVWVSAFLSAEVLTDIGATDITKGFPTVNLNFW
NVKSYEGETWLKFDGKQFYADINVLQNPDSVGHVGYEIIYGYKPAHNSGTET
LPVKVQLPDPFYVTLDSIYIENDLPINLAMETWITRKPDQTSVSSGDVEIMVWFYNN

QY 11MPGGQKVDEFTTTIIBINGSVPVETKNDVYFAPKWDYLAFLRTIPMKDGRVKNVKD
FVEKAAEVIKKHSTREVNFDENYFCVWEIGTEFGDENTTAARFGMTKDFSFVEIGE"
ORIGIN
Query Match 9.3%; Score 89.2; DB 1; Length 2803;
Best Local Similarity 49.8%; Pred. No. 1.5e-13;
Matches 325; Conservative 0; Mismatches 303; Indels 24; Gaps 3;

280 GAGATGACGTACAATTAACAGCGCGCTCTACTACGTCCCAACAACCTTGACAACATT 339
Db 924 GAAACGTGGCTCAAGTTCGATGGTGAAGAGGTCCAGTTCTACGGGATATCTACAACATC 983
QY 340 GTCTTGAGGGATAGAAGTAATTTGGTGATGATACCCGGAATATCTATGGAACAAG 399
Db 984 GTTCTTCAAAACCCGGACAGCTGGGTTCAGGCTATCCAGAGATATATACGGTACAG 1043
QY 400 CCATGGAATGCAAACTACGCAACTGATGCCCAATACCATTTACCCAGTAAAGTTTCAAA 459
Db 1044 CCGTGGCGGCGCACAAC-----AGCGGAACAGAGATTTCTCCGGTGAAGTGAAGAT 1097
QY 460 CTAACAGACTTCTATCTAACATCTCTTAACTTGAGCCCAAGAGCGCTGCCAATT 519
Db 1098 CTTCTGTACTTTTACGTAAACCTTGATTTCAATCTGTGTACGGAACGACCTTCCGATC 1157
QY 520 AACTTCGCAATAGAATCTCTGTTAAACGAGAGAAGCTTGGAGAACAACAGGAATTAACAGC 579
Db 1158 AACCTTGCATGGAACCGTGGATCAACAAGAACGAGATCAGACATCTG---TTTCTCT 1214
QY 580 GATGACGAAGAGTAATGATATGATGATTTACTATGAGGATTAACACCGGTGGCTCCAAA 639
Db 1215 GGAGACGTGAGATCATGCTCTGTTCTTACAACAACATCTCTGATGCCAGGTGGCAGAAG 1274
QY 640 GTTAAAGGAGATGTAGTCCCAATATAGTTAAACGGAACACAGTAAATGCTACATTTGAA 699
Db 1275 GTGGATGATTTCAACAACACGATCGAGATAAACGGCTCTCTGTGGAACAANAATGGAC 1334
QY 700 GTATGGAAGGCAACATTTGGTTGGAGATGATGTGATTTAGATTAAGAACCCCAATCAA 759
Db 1335 GTTTACTTTTGGCCCTCGGGGCTGGGATTTATCTCGCTTCAGGCTCAACAACCTCCATGAA 1394
QY 760 GAGGGAACAGTGACAATTTCCATACGAGCATTTTAAAGTTGTGAGCCCAACATTT---- 814
Db 1395 GATGCAGGGTGAAGTTCAACGTGAAGGACTTTGTGGAAGAGCGGCGGAAGATTATCAA 1454
QY 815 -----CAAGCTTACCAAAATTACAGAACTTTTACTTAGAGAGCTGGAGATTGGA 864
Db 1455 AAACACTCAACAAGGGTGGAAAACTTCGATGAGATGTACTTCTGCTCTGGGAGATAGGA 1514
QY 865 ACTGAGTTTGGACCGCAGACACTACCTCCGCCACCTAGAGTGGTGATCA 916
Db 1515 ACAGAGTTCCGTGATCCAAATACCACCGCGCAAAATTTGGATGAGCACTTCA 1566

RESULT 9
TNCEL A
LOCUS T.maritima celA and celB gene.
DEFINITION T.maritima celA and celB gene.
ACCESSION Z69341
VERSION Z69341.1 GI:1297060
KEYWORDS celA gene; celB gene; cellulase; endo-1,4-beta-glucanase.
SOURCE Thermotoga maritima
ORGANISM Thermotoga maritima
REFERENCE 1 (bases 1 to 2800)
AUTHORS Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
TITLE Liebl, W., Ruile, S., Bronnenmeier, K., Riedel, K., Lottspeich, F. and
Liebl, W.
JOURNAL Microbiology 142, 2532-2542 (1996)
REFERENCE 2 (bases 1 to 2800)
AUTHORS Liebl, W.
TITLE Direct Submission

```

JOURNAL Submitted (02-FEB-1996) Liebl W., Technische Universitaet Muenchen,  
Lehrstuhl f. Mikrobiologie, Arcisstr. 21, 80290 Muenchen, Federal  
Republic of Germany

## FEATURES

source 1..2800  
/organism="Thermotoga maritima"  
/mol\_type="genomic DNA"  
/strain="MSB8 (DSM3109)"  
/db\_xref="taxon:2336"  
406..968  
repeat\_region  
/citation=[1]  
/rpt\_type=OTHER  
/rpt\_unit="406..435"  
1091..1101  
/citation=[1]  
1107..1880  
/gene="celA"  
1107..1880  
/gene="celA"  
/EC\_number="3.2.1.4"  
/function="hydrolysis of carboxymethyl cellulose,  
beta-glucan, etc."  
/standard\_name="cellulase"  
/citation=[1]  
/codon\_start=1  
/evidence=experimental  
/transl\_table=1  
/product="endo-1,4-beta-glucanase"  
/protein\_id="CAA93273.1"  
/db\_xref="GI:1297061"  
/db\_xref="GOA:Q60032"  
/db\_xref="SPTREMBL:Q60032"  
/translation="MVLMTKPGTSDFTVWNGIPLSMELNLWNIKEYSGSVAMKPDGKIK  
TFDADONLSPKPERVLCVPRFYGYKPDWENHTAEGSKLPVPSMKSFSEVVSFD  
IHPEPSLPINMETWLTREKYQTEASIGDVEIMWFYNNLTPGGEKLEETIPFVL  
NGSEVGTWELWLAEDWYLAFLKDFPVKGRKEDVRFHFLDAAGLAUSSARVKDF  
EDLYFTWEIGTEFGSPETKSAQFGWKFNFSIDLEVRE"  
1862..1872  
/gene="celA"  
/citation=[1]  
1877..2701  
/gene="celB"  
1877..2701  
/gene="celB"  
/EC\_number="3.2.1.4"  
/function="hydrolysis of carboxymethyl cellulose,  
beta-glucan, etc."  
/standard\_name="cellulase"  
/citation=[1]  
/codon\_start=1  
/evidence=experimental  
/transl\_table=1  
/product="endo-1,4-beta-glucanase"  
/protein\_id="CAA93274.1"  
/db\_xref="GI:1297062"  
/db\_xref="GOA:Q60033"  
/db\_xref="SPTREMBL:Q60033"  
/translation="MRWAVLLMVVFSALLFSSEVLTSGVATDIFNGFPVTVMLNFW  
NVKSYEGETWKFGEKVEFYADILNVLQNSDWSVHGYPEIYGYKPDWAGHNSGVF  
LPVKGLDPPFYVTLDSYIWNENLPLINLAMEWTITRSPDQTSVSGDAEIMWFVYNN  
VLPGGQKQVEFTVTVEINGVKQETWDVYFAPKWDYLAFLRTTPMKGEKVKINVD  
FVQAAVVKKHSTRINLFEILFCWKEIGTEFGDPNTTAAKGTWTFRFSVEVWK"  
1877..1927  
/gene="celB"  
/citation=[1]  
1928..2698  
/gene="celB"  
/product="endo-1,4-beta-glucanase"  
2722..2744  
/citation=[1]

## ORIGIN

Query Match 8.8%; Score 84.2; DB 1; Length 2800;

Best Local Similarity 50.2%; Pred. No. 3.6e-12;  
Matches 305; Conservative 0; Mismatches 278; Indels 24; Gaps 3;  
QY 332 ACAACATTGTCCTTGAGGATAGAAGTAATTTGGTGCATGGATACCCCGAATATTCTATG 391  
|||||  
Db 2082 ACAACATCGTTCTTCAGAAATCCAGACAGCTGGTGCATGGATATCCAGATCTACTACG 2141  
QY 392 GAAACAAGCCATGGAATGCAAACTAGCAACTGATGGCCCAATACCATTCACCAAGAA 451  
|||||  
Db 2142 GTTACAAGCCCTGGCGGGGCACAAACAGCGGTGTGAAATTTCTCTCTGTG-----AAGG 2195  
QY 452 TTTCAAACCTAACAGACTTCTATCTAAACAATCTCTATAAACTTCGAGCCCAAGAACGGCC 511  
|||||  
Db 2196 TGAAGATCTTCGGGATTCTACGTGACTCTTGATTTACTCGATCTGGTACGAACAAC 2255  
QY 512 TGCCAAATTAACTTCGCAATAGAAATCCTGGTTAAACAGAGAGAGCTTGGAGAACCAACAGGAA 571  
|||||  
Db 2256 TGCCTATCAACCTTGCCTATGGAGAGCTGGATCACAGAAAGTCCCGACCAGACCTCTG--- 2312  
QY 572 TTAAACAGGATGAGCAAGAAGTAATGATATGATTTACTATGACGGATTACAAACCGGCTG 631  
|||||  
Db 2313 TTTCTTCGGTGACCGGGAGATCATGTTTGGTTCTCAACAACGTTCTGATGCCGCG 2372  
QY 632 GCTCCAAAGTTAAGGAGATTGTAGTCCCAATAANTAGTTAACGGAACACACAGTAAATGTA 691  
|||||  
Db 2373 GTCAGAAAGTGGACGAGTTCACACACAGTCGAGATAAAGCGGCTGAAGCAGGAAACAA 2432  
QY 692 CATTTGAAGTATGGAAGGCAACATTTGTTGGGATGATGTTGCATTTAGATAAAGACCC 751  
|||||  
Db 2433 AGTGGGATGTTTACTTTGACCGTGGGATGGGATTTACTCTTTCAGCTGACAAAC 2492  
QY 752 CAATCAAGAGGGAACAGTGCATTTCCATACGGAGCAATTTATAAGTGTTCGAGCCCAACA 811  
Db 2493 CGATGAAGAAGAAAGGTGAAAAACAAGTGAAGGACTTCGTTCAGAAAAGCCGCGAAG 2552  
QY 812 TT-----TCAAGCTTACCAATTACAGAGACTTTTACTTAGAGGCGGTGG 856  
Db 2553 TTGTCAAGAGCACTCAACAGAGATAGACAAATTCGAAGAGCTGATTTCTCGCTGTGG 2612  
QY 857 AGATTGGAACCTGAGTTTGGAAACGCCAAGCACTTACCTCCGCCCCACCTAGAGTGGTGATCA 916  
|||||  
Db 2613 AGATCGGAACGGAGTTTGGAGATCCGACACACACAGCGCGGAAATTCGCTGGACTTCA 2672  
QY 917 CAAACAT 923  
Db 2673 GGGACTT 2679

## RESULT 10

AE001800 20653 bp DNA linear BCT 02-JUN-1999  
LOCUS Thermotoga maritima section 112 of 136 of the complete genome.  
DEFINITION Thermotoga maritima section 112 of 136 of the complete genome.  
ACCESSION AE001800 AE000512  
VERSION AE001800.1 GI:4982090  
KEYWORDS  
SOURCE Thermotoga maritima  
ORGANISM Thermotoga maritima  
Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
REFERENCE 1 (bases 1 to 20653)  
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,  
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,  
McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,  
Stewart,A.W., Cotton,M.D., Pratt,M.S., Phillips,C.A.,  
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,  
Eisen,J.A., Fraser,C.M. et al.  
EVIDENCE for lateral gene transfer between Archaea and bacteria  
from genome sequence of Thermotoga maritima  
Nature 399 (6734), 323-329 (1999)  
TITLE  
JOURNAL  
MEDLINE 98287316  
PubMed 10360571  
REFERENCE 2 (bases 1 to 20653)  
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,  
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,



McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,  
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,  
Richardson, D., Reidelberg, J., Sutton, G.G., Fleischmann, R.D.,  
White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

## FEATURES

Location/Qualifiers

1..20653  
/organism="Thermotoga maritima"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:2336"  
840..1616  
/gene="TM1524"  
840..1616  
/gene="TM1524"  
/note="similar to PID:1297061 GB:AE000512 percent  
identity: 100.00; identified by sequence similarity;  
putative"

/codon\_start=1  
/transl\_table=11  
/product="endoglucanase"  
/protein\_id="AAD36591.1"  
/db\_xref="GI:4982091"  
/translation="MVMIMTKPGTSDFWNGIPLSMELNWNKEYSGSVAMKFDGEX  
ITPDADIQNSPKPEPVYLGYPFYGYKFWENHTAEGSKLPVPSMKSFSVSEVF  
DIHEPSLPLNFAWETLREKYQTEASIGDVEIMVMFYFNLTGCGEKIEETIPFV  
LNGESVGTWLAEBWGDYLAFLKDPVKGRKVPDVRHFLDAAGALLSSARVKD  
FEDLYFTWETGTFEGSPETKSAQFGNKFENFSIDLEVR"  
1613..2437  
/gene="TM1525"  
1613..2437  
/gene="TM1525"  
/note="similar to GP:1297062 percent identity: 100.00;  
identified by sequence similarity; putative"

/codon\_start=1  
/transl\_table=11  
/product="endoglucanase"  
/protein\_id="AAD36592.1"  
/db\_xref="GI:4982092"  
/translation="MRWALIMVFSALLFSSEVYLTSGATDISNPGFPVTWELNFW  
NPKSYEGETWLPDGEKVEFYADLVNLPQDSWVHGYPEIYGYKPMAGHNSGVF  
LPVKQDLPDYVTLDYSIWYNNLPINLAMETWITRSPDTSVSSGDAIMVMFYNN  
VLMGQGVQDEFTTVEINGVKQTKWDYFAPKWDYLAFLITPMKEGKVINVD  
FVQAAAVKXKSRIDNFBELFCVWEIGTERGDPNTIAAKFGWTFRDSVEVK"  
2557..2994  
/gene="TM1526"  
2557..2994  
/gene="TM1526"  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAD36593.1"  
/db\_xref="GI:4982093"  
/translation="MKVSDILTVAIRLEERGERFYRLSHFNGEIKTKTLELADOER  
THAIFRMSDOENWSDVSLAGYAFVEFDTSEILREKDLTKVELVDIAISVEKD  
SILLYELKGLNSDAQTKVKKIIDQEKHLKLEMKRES"  
complement(2956..3378)  
/gene="TM1527"  
complement(2956..3378)  
/gene="TM1527"  
/note="similar to GB:AE000657 percent identity: 54.92;  
identified by sequence similarity; putative"

/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36594.1"  
/db\_xref="GI:4982094"  
/translation="MPGCKOKHRYALRLVRLSLNRPMSSELAEKELITRNPAK  
VLHYLRKGLWNSIKGNGGRVLAKSPDEISFLDVISVEKDLVIVDCNSRCKNYRAC  
RVKYFNWLSYLDLFSNTIKDIASGTTFHLEKLS"

gene  
CDS

3406..4275  
/gene="TM1528"  
3406..4275  
/gene="TM1528"  
/note="similar to GB:X73124 SP:P39582 PID:413930  
GB:AU009126 percent identity: 57.75; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="1,4-dihydroxy-2-naphthoate  
octaprenyltransferase, putative"  
/protein\_id="AAD36595.1"  
/db\_xref="GI:4982095"

/translation="MKYFVAVRPFVSASFVTVGAMLAEEFSLIRYILSLIASIFL  
HAGVNTNDYDKKGVTDKDSGLSSGLLVSKITPREELLVSFCVAVSVLGLVLI  
KISGALLWGLTGLVGYAVTGHPFYLYKSLGMLVFLMGLPLVGLVAYVOTGRF  
SLRALLSVIPYGATDILLANEIRDSEFDRSGIKTLPILIGDRAASFVAVTGLI  
YVFVILVSTGVFRVLSLVSLVSLPLYIRVTKQLFQKSAGRNAREIADVDKMSALAE  
MLFVSMILGLLR"  
4275..4901  
/gene="TM1529"  
4275..4901  
/gene="TM1529"  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAD36596.1"  
/db\_xref="GI:4982096"  
/translation="MPMVKRISEISFLSLGLFLGMVYFQVNMDFWMEFMTTLI  
PSVISIFLGGRNKPNKYEAPLAYISVAAYLLFVGYLISLLIPFHRDVSVYGL  
AEGNALKVIIILVIAFFETIWRGFVTEFLQRLDVPVPSLISLLISVSVHFTGN  
MALIVAFVLIIGIISLLYVITGVSTPAFAHALMSLLIFVVLPLKGGF"  
5003..5860  
/gene="TM1530"  
5003..5860  
/gene="TM1530"  
/note="similar to PID:1055220 SP:P52040 percent identity:  
66.53; identified by sequence similarity; putative"

/codon\_start=1  
/transl\_table=11  
/product="electron transfer flavoprotein, beta subunit"  
/protein\_id="AAD36597.1"  
/db\_xref="GI:4982097"  
/translation="MNVVVCIKQVPTNTVIRIDRKNLVREGVPSIINPDRALEL  
ASQKEKFGATVYITMGPPQAKKADAIAPGLDEAVHLSDRTPAGADTLATYTL  
WGIKIEERIIGKIDILITGQAVDGTQGVGLATPFGYALGAYVVRIBEDPEKKE  
MVIVRLDQGFELRLKLPVLTITDELNKPAYADLPNLRIRIYEPVITWTKDLGLD  
PKKCGFGSPTRVVTNIPPAKGGDIISKNEDEPVAEKLEALKKKFEAVRLVEALK  
PVLEGEKDE"  
5853..6869  
/gene="TM1531"  
5853..6869  
/gene="TM1531"

/note="similar to PID:1055221 SP:P52039 percent identity:  
58.15; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="electron transfer flavoprotein, alpha subunit"  
/protein\_id="AAD36598.1"  
/db\_xref="GI:4982098"

/translation="MSEKKIFVLIEHHGKAHPVSWELIGKARDLASKLSENVWGV  
LIGELGESVAEAIQRGADKVLVYKQNRNFNTVYVLYKALVDVVRKYRPEIFLIGAT  
LEGRLAGWATELETGLTDIIPDKKLLMATPTFGNLMATIMCPDRPQM  
ATVRGVVKELPPDPERTGEIIEEYDGLTKLIEILETIPLOTQVNLVAVPVQAG  
CKGVGPGEPFKLLELADLLGGEVGAASAUKAGWISPEHOVSGTKTVRVLVYACG  
ISGALQHVVGIGSEIIVAINIDEKAPFDIADIGVGDLLKVVPAITAKLRELINKS  
GVKK"  
6866..8182  
/gene="TM1532"  
6866..8182  
/gene="TM1532"

gene  
CDS

3406..4275  
/gene="TM1528"  
3406..4275  
/gene="TM1528"  
/note="similar to GB:X73124 SP:P39582 PID:413930  
GB:AU009126 percent identity: 57.75; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="1,4-dihydroxy-2-naphthoate  
octaprenyltransferase, putative"  
/protein\_id="AAD36595.1"  
/db\_xref="GI:4982095"

gene  
CDS

3406..4275  
/gene="TM1528"  
3406..4275  
/gene="TM1528"  
/note="similar to GB:X73124 SP:P39582 PID:413930  
GB:AU009126 percent identity: 57.75; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="1,4-dihydroxy-2-naphthoate  
octaprenyltransferase, putative"  
/protein\_id="AAD36595.1"  
/db\_xref="GI:4982095"



/note="similar to SP:P26484 GB:X55450 PID:38698 percent identity: 66.82; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="fixC protein"  
/protein\_id="AAB36599.1"  
/db\_xref="GI:4982099"  
/translation="MKIEFDVVVVGAGPSGLSAYVLAKNGLKVAIVVEKEVPGSQNV MGVLVYHPLEKEIMPDPLEKARNVIEQNLWLLGNEGVIKIGHNVWEKKNP NAFVLRANPDRLFAQVEKAGALIIPTKVEDFLANEKEIAGVTSFPGKBIHSA VLAEGVNPILTWKAGRKEDLKPVMVAIVKEVISVPEDVNRVFGVGNDAATIEL LGSSEGMFGFGLPYANESVSLGCGVLLEDLKKIKPKVQLLENKHPVLSMDLGE YRNIMYELHPLIPEGGIYAMPKYGDRVLVCGDAAMLVNSIHREGSNHAIISGRLLA ETLEAEKGFDFEKLKYLRLKESFILKLEKYDLMTPEKKNHQFVEIYFDLAN DALKRFVQDGTGPKWDQKQIADWLSRSLIGISLDLRFWEAV"  
8189.8467  
/gene="TM1533"  
8189.8467  
/gene="TM1533"  
/note="similar to GB:X65515 PID:510488 SP:P53658 percent identity: 68.82; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ferredoxin"  
/protein\_id="AAB36600.1"  
/db\_xref="GI:4982100"  
/translation="WRLEDKLYLNRYTDBENPHLKIKBSICAEKCSRRPCVSCCPA DVEWTSSEMEVFEGLCEGTCRIVCPFGNIENYPRNGYLVLYKFG"  
8477.8818  
gene  
CDS  
Query Match 8.8%; Score 84.2; DB 1; Length 20653;  
Best Local Similarity 50.2%; Pred No. 3.7e-12;  
Matches 305; Conservative 0; Mismatches 278; Indels 24; Gaps 3;  
Qy 332 ACAACATTGCTTGAGGATAGAAAGTAATTGGTGCATGGATACCCCGAATATTCTATG 391  
Db 1818 ACAACATGTTCTTCAGAAATCCACAGCTGGGTGCATGGATATCCAGAGATCTACTACG 1877  
Qy 392 GAACACAGCCATGAATGCAACATCCGAACTGATGCCCAATACCATTTACCAGTAAG 451  
Db 1878 GTTACAAGCCCTGGGGGGGACACACAGCGGTGTGGAATTCTTCCTGTG-----AAGG 1931  
Qy 452 TTTCAAACCTAACAGACTTCTATCTAACAATCTCCTATAAACTTGAGCCCAAGAACGCC 511  
Db 1932 TGAAGATCTTCGGAATTTACGTGACTCTTGATTACTCGATCTGTGAGAAACAAC 1991  
Qy 512 TGCCAAATTAATTCGCAATAGAAATCTTGTTAACGAGAGAGCTTGAGAACACAGGAA 571  
Db 1992 TGCCTATCAACCTTGCCATGAGAGCTGGATCAGAGAAGTCCCGACGACCTCTG--- 2048  
Qy 572 TTAACAGCGATGACAGAACTATGATATGATTTACTATGACGGATTACAACCGGCTG 631  
Db 2049 TTTCTTGGGTGACGGGAGATCATGGTTGGTTCTTACACACAGTTCTGATGCCGGCG 2108  
Qy 632 GCTCCAAAGTTAAGGAGATTGTAGTCCCAATAATAGTTAAGCGAACACCAAGTAATGCTA 691  
Db 2109 GTCAGAAAGTGCAGAGTTCACTACACAGTCGAGATAAAGCGGCTGAAGCAGGAAACAA 2168  
Qy 692 CATTTGAGTATGGAAGCAACATTTGGTTGGAGTATGTTGCAATTTAGATAAAGACCC 751  
Db 2169 AGTGGGATGTTTACTTTTGACCGGTGGGATGGGATTAACCTTGCTTTTCACTGACACAC 2228  
Qy 752 CAATCAAGAGAGGAAACAGTACAATTCATACCGGAGCATTTTAAAGTGTTCAGCCCAACA 811  
Db 2229 CGATGAAGAGAAAGAGTGAATCAACGTTGAAGAGCTTCGTTCAAGAAAGCCCGGAAG 2288  
Qy 812 TT-----TCAAGCTTACCAATTAACAGAACTTTACTTAGAGACCTGG 856  
Db 2289 TTGTCAAGAGACACTCAACAGAGAAATAGACAAATTTTCAAGAGCTGTATTTCTGGCTCTGGG 2348  
Qy 857 AGATTGGAAGTGGTTTGGACCGCCAGCACTACTCTCGGCCCACTAGAGTGGTGATCA 916

Db 2349 AGATCGGAACGAGTTTGAGATCCGACACACACGCGGAAATTCGCTGGACTTCA 2408  
Qy 917 CAACAT 923  
Db 2409 GGGACTT 2415  
RESULT 11  
TN286103 4547 bp DNA linear BCT 22-SEP-1997  
LOCUS T.neapolitana cela and celB genes and two open reading frames.  
DEFINITION Z86103  
ACCESSION Z86103  
VERSION Z86103.1 GI:1870178  
KEYWORDS cela gene; celB gene; cellulase; endo-1,4-beta-glucanase.  
SOURCE Thermotoga neapolitana  
ORGANISM Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
REFERENCE 1  
AUTHORS Dakhova,O.N., Kurepina,N.E., Zverlov,V.V., Svetlichnyi,V.A. and Velikodvorskaya,G.A.  
TITLE Cloning and expression in Escherichia coli of Thermotoga neapolitana genes coding for enzymes of carbohydrate substrate degradation  
JOURNAL Biochem. Biophys. Res. Commun. 194 (3), 1359-1364 (1993)  
MEDLINE 93356813  
PubMed 8352795  
REMARK (sites)  
REFERENCE 2 (bases 1 to 4547)  
AUTHORS Zverlov,V.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 4547)  
AUTHORS Zverlov,V.  
TITLE Direct Submission  
JOURNAL Submitted (27-FEB-1997) Zverlov V., Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Sq., Moscow, Russia, 123182  
FEATURES  
source  
1. .4547  
/organism="Thermotoga neapolitana"  
/mol\_type="genomic DNA"  
/strain="Z2706-MC24"  
/db\_xref="taxon:2337"  
1. .326  
/gene="orf1"  
/pseudo  
1. .326  
/gene="orf1"  
/pseudo  
/codon\_start=3  
/transl\_table=11  
/db\_xref="PSEUDO:CAB06780.1"  
756.761  
/db\_xref="RENTREMBL:CAB06780"  
767.2332  
/gene="orf2"  
767.2332  
/gene="orf2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAB06781.1"  
/db\_xref="GI:1870179"  
/db\_xref="STREMBL:P96490"  
/translation="MKSIEGLKTEPKYAFSLPANYMLLENLPKECEVLCILKLA WGNFKBAIQTKILKSKGEAAKFLANKLAFKYTKGVDMMLYLXKNLPMKS KSIQRTIVITLINFASGKPKLRKRVVNDYRKSTLSFLYLARRADSGRLSEAV HDYIOAQLSRVPHPTCTIVSLNDLAWDIRKPKLAHALSQGAFVWLYVREPGN LSGADLTLPVKEKMDSPSIHSTAHIVSLPVPEDYLSLKKAKFVLDYTCSTYPTN SQRLRYVEKAWKGTLSKGLSDILKGTWIRADTIRKLTJSCVDTCAPFPVNEW IKMERLKYRESSEKLPISLSEINPFPTTYTLYALLDREFLSRERLKRKYLIEDI ESFADFMADKHETMEFVSMVKAHPFIEGRKETVRKALGRMKRKLRTLLRIENKE SDRKLLDRFLNRYGDVRFVGRVLRKGPDWVREFARRYSKVOPLLVAFWCSEDRVR

Db 3767 AGATCATGCTGGTCTTATTTCAACGAACCTCACACCGGTCGAAAGAAAGTGGAGAGT 3826  
 QY 650 TTGTAGTCCCAATAATAGTTTACGGGAACACACAGTAAATGCTACATTTGAAGTATGGAAGG 709  
 Db 3827 ACACCGTTTCGTTGGAACCTTAAACGGGAACAAACAAACGTTGGAATCTGGGAATCTCTGGCAG 3886  
 QY 710 CAAACATTTGGTTGGGAGTATGTTGCAATTTAGAAATAAGACCCCAATCAAAAGAGGGAACAG 769  
 Db 3887 CCGAATGGAACCTGGGATTTATCTGTTCCGCTCAAGAAATCCTGTGAAAGAGGAGGG 3946  
 QY 770 TGACAATTCATACGGAGCAATTAAGTTGTCGACCAACATTT-----CAA 817  
 Db 3947 TGAGATTCAACGTGAAGGATTTTCTCGATGTTGCGAGGAGATCTTTCAAGTCCACAC 4006  
 QY 818 GCTTACCAAAATTACACAGAACTTTACTTTAGAGAGCTGGAGATTTGGAATCTGAGTTTGGAA 877  
 Db 4007 GTGTGAAGGATTTGATGATCTGATTTTACCGTCTGGAGATCGAACCAGATTTGGA 4066  
 QY 878 CGCCAAGCACTACCTCCGCCACCTAGAGTGTGTTGATGATCAAAACATAACACT 929  
 Db 4067 GGCCTGAAACAAAGAGTGCCTGTTTGGATGACGCTTCAATAACTTTTCAAT 4118

RESULT 12  
 166494/c  
 LOCUS 166494 7218 bp DNA linear PAT 28-DEC-1997  
 DEFINITION Sequence 14 from patent US 5670367.  
 ACCESSION 166494  
 VERSION 166494.1 GI:2724471  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 7218)  
 AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F., Gunter.  
 TITLE Recombinant fowlpox virus  
 JOURNAL Patent: US 5670367-A 14 23-SEP-1997;  
 FEATURES  
 source  
 1..7218  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

ORIGIN  
 Query Match 5.6%; Score 53.8; DB 6; Length 7218;  
 Best Local Similarity 5.4%; Pred. No. 0.00099;  
 Matches 22; Conservative 218; Mismatches 165; Indels 0; Gaps 0;

QY 502 AAGAACGGCTCCCAATTAACCTCGCAATAGAACTCTGTTAAACGAGAGAGCTTGGAGA 561  
 Db 1447 AAGAATTTGGTACRR 1388  
 QY 562 ACAACAGGATTAACAGCGATGAGCAAGAGTAATGATATGATTTACTATGACGATTA 621  
 Db 1387 RRR 1328  
 QY 622 CAACCGCTGGCTCCCAATTAAGGAGATTTAGTCCCAATAAGTATTAACGGAACCA 681  
 Db 1327 RRR 1268  
 QY 682 GTAATGCTACATTTGAAGTATGGAAGCAACATTTGTTGGAGTATGTTGCATTTAGA 741  
 Db 1267 RRR 1208  
 QY 742 ATAAAGACCCCAATCAAGAGGAAAGTGAACATTCATACGAGAGCTTTTAAAGTGT 801  
 Db 1207 RRR 1148  
 QY 802 GCAGCAACATTTCAAGCTTACCAATTAACAGAACTTTTACTTAGAGAGCTGGAGATT 861  
 Db 1147 RRR 1088  
 QY 862 GGAATGATGTTGGAAACCCCAAGCACTACCTCCGCCACCTAGAG 906

RRLTILRYMVG"  
 3354..3358  
 3367..4140  
 /gene="celA"  
 3367..4140  
 /gene="celA"  
 /EC number="3.2.1.4"  
 /function="hydrolysis of beta-1,4-glucoside bonds"  
 /standard\_name="cellulase"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="endo-1,4-beta-glucanase"  
 /protein\_id="CA306783.1"  
 /db\_xref="GI:1870180"  
 /db\_xref="GOA:P96491"  
 /db\_xref="SPTREMBL:P96491"  
 /translation="MVELTAPGTADFWRNDMPLSMELNWNVRYTGTVMRFDGERL  
 TFNGDVELSAREPERVILGPEFYGYKPEWHAAGTKLPLVSVESGFTVELSEFL  
 IDHPSLPLNFAETWTREKYQVEASIGDVEIMVWFNPELTPGGRKKVGEYTVSPEL  
 NGEQRGIWEIWHAEWMDYLAFRLKNPVRKGRVRFNVKDFLDVAGEYLSRSTRVKDF  
 DDLYFTVMEIGTERGSPETKSARFGWTFNNFSDIMEVKG"  
 4125..4129  
 /gene="celA"  
 4137..4547  
 /gene="celB"  
 4137..4547  
 /gene="celB"  
 /EC number="3.2.1.4"  
 /function="hydrolysis of beta-1,4-glucoside bonds"  
 /standard\_name="cellulase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="endo-1,4-beta-glucanase"  
 /protein\_id="CA306782.1"  
 /db\_xref="GI:1870181"  
 /db\_xref="GOA:P96492"  
 /db\_xref="SPTREMBL:P96492"  
 /translation="MRILVSLVLSAFPLSAEVLVDIGATDITFKGPFVWMLNFW  
 NVKSYGBETWIKPQKQVYADYINVLQNPDSWVHGYPEIYGYKPMWAHNSGTET  
 LPVKVLPDYFVTLDSYIWNELINLAMETWI"

ORIGIN  
 Query Match 8.5%; Score 82; DB 1; Length 4547;  
 Best Local Similarity 49.2%; Pred. No. 1.5e-11;  
 Matches 350; Conservative 0; Mismatches 335; Indels 27; Gaps 4;

QY 230 TCTACATGAATAAACCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGT 289  
 Db 3422 TTTCATGGAGCTGAATCTGTGGAACGTAGAAAGATACACGGGAACGGTGTGATGAGAT 3481  
 QY 290 ACAATTTAACAGCGGCTCTTCACTACGTCCCAACACTTGACAACTTGTCTTGAGGG 349  
 Db 3482 TCGA-----CGGTGAAGAGCTACCTTCAACGGTGATGTTGAAGATCTTTCTGCCAGAG 3535  
 QY 350 ATAGAAGTAATGGGTGATGATACCCCGAAATATCTATGGAACCAAGCCATGGAATG 409  
 Db 3536 AGCCAGAAGGTACATCTTGGCTATCTGAGTTTACTACGATACAAACCTGGGAAA 3595  
 QY 410 CAACACTACGAACTGATGCGCCCAATACATTTACCAGTAAGTTTCAACCTACAGACT 469  
 Db 3596 GACATGACGACAGAG-----GGACAAAGTTTACCTCTGTTAGTCTTCTGTTGAGTCCT 3649  
 QY 470 TCTATCTAACCAATCTCTATATACTTGAAGCCCAAGAACCGGCTCCCAATTAACCTGCCAA 529  
 Db 3650 TCACGGTCAACTTCTTCCGAATAGACACATGCGCTCTCTTCCGCTGACTTGGCA 3709  
 QY 530 TAGAATCTCTGTTAAACGAGAGAGCTTGGAGAAACACAGGAATTAACAGCGATGAGCAAG 589  
 Db 3710 TGGAAACCTGCTGCACAAAGAGAAAAATACCAGGTGGAAGCTTCCATCGGAGATGTG---G 3766  
 QY 590 AAGTAATGATATGATTTACTATACCGGATTTACAAACCGGCTCGCTCCCAAGTTTAAGAGA 649

Db 1087 RRRRRRRRRRRRRRRRRRRRRATCGAAGTCCCTCGACCTGCAG 1043

RESULT 13  
 AX655393/c  
 LOCUS AX655393 2000 bp DNA linear PAT 22-MAR-2003  
 DEFINITION Sequence 5263 from Patent WO03000898.  
 ACCESSION AX655393  
 VERSION AX655393.1 GI:29158207  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 ORGANISM  
 Oryza sativa  
 Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhacordeae; Oryzaceae; Oryza.

REFERENCE  
 1 Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,  
 Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.  
 Plant genes involved in defense against pathogens  
 TITLE Patent: WO 03000898-A 5263 03-JAN-2003;  
 JOURNAL Syngenta Participations AG (CH)  
 FEATURES  
 Location/Qualifiers  
 source  
 1..2000  
 /organism="Oryza sativa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:4530"

ORIGIN

Query Match 4.9%; Score 47.4; DB 6; Length 2000;  
 Best Local Similarity 10.4%; Pred. No. 0.057;  
 Matches 75; Conservative 318; Mismatches 324; Indels 4; Gaps 2

QY 10 AAAAAGTTGCTGATGATCTATCTTACCAATCCTTTAGTACAGGCAATATATTGTA 69  
 Db 944 RHEAWCYCCKWKWKWTSCKWKYRTHSWCHWTWGWAMRYAYAMRRRTYKWSR 885  
 QY 70 GAAAAGTATCATCCTCTGAGGACAAGTCACCTTCAAATACCTCATCTACACACCCAA 129  
 Db 884 RMYTWTWKAWTMTCMCAKXWYATGWATWMMWRYTMYCYAMTCAKCYKAMTKWM 825  
 QY 130 ACAACACITTCACACTACCAAGTTCTCAAGATATAGATACCTGATGACGCTGAGTGCCA 189  
 Db 824 TTWACHAATSRATSRWAMAGMKWKRYKMKRATYMWKRWRCWAGWARKWKRYKWKY -- 767  
 QY 190 GGAGCTCCTATTGATAAGGATGGTGATGGGAACCCAGAATTTACATTGAAATAAACCTTA 249  
 Db 766 ATRYKXWAMTMMWSNRKWSYRMMWSGMRMWSAWRYCSRMKCAKTKYASSARWTKRA 707  
 QY 250 TGGACATTCITTAAGTCTACTGATTTGCTGAGATGACGTACATTTTACCAGCGCGCTC 309  
 Db 706 KRSYRRRWYKRGWYTRYRYRSCMTFARMKSKRWAGASMKSCMYWRGASMMY 647  
 QY 310 CTTCACTAC - GTCCAAACAATTCGACAACTTGCTTGGGAGATAGAAGTAATTTGGGTGC 367  
 Db 646 SKYSCSKCKKTRPMTSSYMTSGYMYSSYKSWTSKWSYMGKMTCTWYTSMKGSTR 587  
 QY 368 ATGAGTACCCGGAATATTCTATGAAACAAGCATGGAATGCAACTAGCAACTGATG 427  
 Db 586 RSKMGWSGSRMYRMRWKKRKRKYRMYKWKCKTWRRCMYRGWYTTWTSRSMYTG 527  
 QY 428 GCCCAATACCATTCACCAAGTAAAGTTTCAAACCTTACAGACTTCTATCTAAACAATCTCT 487  
 Db 526 RYKARYTSKRPYMYKYRYCWYYGYMYMKSYMYRYGYCKACKCCYAMCWKAYSGM 467  
 QY 488 ATAAACTTGAGCCCAAGACCGCCTGGCAATTAACCTTCGCAATAGAAATCTCGTTAAACA 547  
 Db 466 MMYWYKYSKWVRMSTKYMMWSMMYKCRSMKYAKGCGYCKMMWTSYSGYMKWYTYMGSY 407  
 QY 548 GACAGCTTGGAGACAACAAGGAATTAACACCGATGACCAAGAAGTAATCATATGATTT 607  
 Db 406 KYSRCYKYNRMTYKGMWMMYYSAYSSMTWYTYKWKYKRGKTSWTKGYSYKCK 347  
 QY 608 ACTATGACGGATTACAACGGCTGCTCCAAAGTTAAAGGAGATTGTAGTCCCAATAATAG 667

```

Db      346 YCTWVCYMKCMCRYWRKMKRKTYSKRYCWRYATCYCCCYRKGWYSRRSMRTAG 287
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      668 TTAACGGAAACACAGTAAATGCTACATCTTGAAGTATGGAAGCGGAAAACATGGTGGGAGT 727
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db      286 KWKMSWSRWCRSYSWYKMYKWKWKWSYWSYGWARSSTGWSRSAKRTYKGYSTSRRAK 227
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Qy      728 A 728
      :
Db      226 M 226

RESULT 14
LOCUS   AE006803              11595 bp      DNA          linear      BCT 05-JUL-2001
DEFINITION Sulfolobus solfataricus section 162 of 272 of the complete genome.
ACCESSION AE006803 AE006841
VERSION   AE006803.1 GI:13815222
KEYWORDS
SOURCE   Sulfolobus solfataricus
          Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
          Sulfolobus.
REFERENCE
  1 (bases 1 to 11595)
    She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G.,
    Wayes, M.J., Chan-Weicher, C.C., Clausen, I.G., Curtis, B.A., De
    Moers, A., Esaus, G., Fletcher, C., Gordon, P.M., Heikamp, De Jong, I.,
    Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P.,
    Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L.,
    Doolittle, W.F., Duguet, M., Gaasterland, J., Garrett, R.A.,
    Ragan, M.A., Senses, C.W. and Van Der Oost, J.
    The complete genome of the crenarchaeon Sulfolobus solfataricus P2
    Proc. Natl. Acad. Sci. U.S.A. 98 (14), 7835-7840 (2001)
    21332296
    11427726
  2 (bases 1 to 11595)
    She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G.,
    Wayes, M.J., Chan-Weicher, C.C.-Y., Clausen, I.G., Curtis, B.A., De
    Moers, A., Esaus, G., Fletcher, C., Gordon, P.M.K., Heikamp-de
    Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X.,
    Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N.,
    Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, J.,
    Garrett, R.A., Ragan, M.A., Senses, C.W. and Van der Oost, J.
    Direct Submission
    Submitted (24-APR-2001) Europe/Canada joint project: Copenhagen
    University, Denmark; Universite Paris-Sud, France; Wageningen
    University, The Netherlands; Institute for Marine Biosciences &
    University of Ottawa, Canada
    FEATURES
      source
        /organism="Sulfolobus solfataricus"
        /mol_type="genomic DNA"
        /db_xref="taxon:2287"
        complement(219..2204)
        /gene="SSO1948"
        complement(219..2204)
        /gene="SSO1948"
        /note="Multicopy. Similar to SSO1353"
        /codon_start=1
        /transl_table=11
        /product="Hypothetical protein"
        /protein_id="AAK42141.1"
        /db_xref="GI:13815223"
        /translation="MKYTDKDRIVACVPLGGIGTGKLEIDNKKVRIINTVIRNNWGNP
        IKLGRHPVFKDNRKGFIFQDKGLYKVEPGEIITYEGKIPVVKVKGKSEVEE
        LEAFPTIPNDLKNLSLPATGILKVGKIGKVEAFSPNIVGSVSGVNRVESIRNG
        IRKILANDYDPAGKNTLLISNVSDIIITQNLKRKFSSETMFWLSQYNEEPWVKL
        RGEIEDNPEVTHGRDDPASIIISGEEMRYVPFWYFNGKHVPYFGHYENPFDDK
        SETAKYFLDNFDHRLQIPHNIVNVKEWRDALINSYILSSNLTWLDKGRPAIYEA
        PQNCPYLGTGTGYEGSLVPLMWPELEKFLKLIISVYNDGYVPHDILGEHLSDF
        IDGTTSPKKMKNPSLILLYVRYFFXTNDIDELKEVYPTIVKMDHELQCRQDLPT
        MEQMDNAPATIIIGKDSVTSSFTASLAMEIAKLVGDSNTVFGVIRKLNVRGLR
        FRWNGKPAATWDKGVDSYLAQGLGEWFTLLELENIVDENNKALRESIIRLNG

```

```

ASPCVPLNDGKINVLNLSQTYSSWPRLLVFAICWLAYKKGVDLSFCCKEWDNLVR
NGMWQDSRINCCTGPEINYLHDHYVGSFSLASFLE"
/ gene
/ complement (2285. .3269)
/ gene="SSO1949"
/ complement (2265. .3269)
/ gene="SSO1949"
/ EC number="3.2.1.4"
/ note="Energy Metabolism"
/ codon_start=1
/ transl_table=11
/ product="Endoglucanase precursor"
/ protein_id="AAK42142.1"
/ db_xref="GI:13815224"
/ translation="MNKKLYIIIVPIIVGVIGCAIVLHQQSNVKTSTVTT
NETTLMSITVTPTVTPTPTSPOLIVVTSSASPTVLLNNSVPFYLEVNM
NAKTWNGNYWPNPLRTLSVSNLQVNPLOWTNGYBEIYVGRKPMDSYAGNIFP
NRIGNMTPFVSNITKUDPSINFIDASDAKIVRQIAFSPTGAPGNDIEMVWL
FSQNLQPAQQGVGVPIYIYNHTLVNATQVMVKQKVPWGWEYIAFRPDGKWKVNG
YVAYEPNLFIKALNPFASNTNYLTDFEFTGEMTNGTAYFSWTISNYETILL
"
/ complement (3574. .4524)
/ gene="SSO1951"
/ complement (3574. .4524)
/ gene="SSO1951"
/ codon_start=1
/ transl_table=11
/ product="Transposase ISC1190"
/ protein_id="AAK42143.1"
/ db_xref="GI:13815225"
/ translation="MEAPVAGIDVSKDLIVYFQGLYEYENDKRGVEEIRKILPKGC
KVGLESTGVHVNLAKILMGSDYRIINPFILKFKDFGRKSKDNDKAKLAEIVSM
CSGETTDQDFVTRSIARVNRRLRDVLVGRKDSKRNLEEVLRGGDSI
VLAERVFLLELTKRKEIEKENLVSKDSLIETIPICIKTLCIILARVGVK
RFSDKRVAYCGLDPVVSSGKSGVSGISKGDVLRRAFLAALTAIKNPVIRK
FYEBKGLKGLKLVACARKLAVITWAVLYNKPFDASE"
/ 4865. 6045
/ gene="cpsA-2"
/ note="synonym: SSO1952"
/ 4865. .6046
/ gene="cpsA-2"
/ EC number="3.4.17.-"
/ note="Belongs to peptidase family M40, also known as the
ama/hipo/hyc family of hydrolases; Proteases"
/ codon_start=1
/ transl_table=11
/ product="Thermotable carboxypeptidase (cpsA-2)"
/ protein_id="AAK42144.1"
/ db_xref="GI:13815226"
/ translation="MDLVEKLNVDKSIDWIIQIRKIHENPELSYKEYSTGKLVAE
TLRLGIEVEGCVLPVAVGKIRGNKPGKTVALRADMDALPVEETSDVEFKVKGV
MHACGHDTHVAMLGGLYLLVKNKDLISGBIRLIFQPAEDGGLGKAKPMLEAGVNG
VDYFGHISSEYPSGVFATKPFIMATPPAKFIVVHGKGGHSAFHETIDPISLQ
IANAYIGTARIDPQFPFVISITTHSGTKDNIPDDAEMQGTIRSLDENVRKAKO
YMRIVISICIGYATGCEKPMEDVYPTVNNPEVTEVMKILSSISTVVEVPEVLGA
BDFSRFLQKAPGKPFYFLGTRNEKKGCIYPNHSSKFCVDEVDVLKGLAHALLAKFSN
K"
/ 6070. 6549
/ gene="SSO1953"
/ 6070. .6549
/ gene="SSO1953"
/ note="Some similarity with a protein in A. ambivalens sor
5' region (PID: G6002290). Multicopy. Similar to SSO1356"
/ codon_start=1
/ transl_table=11
/ product="Hypothetical protein"
/ protein_id="AAK42145.1"
/ db_xref="GI:13815227"
/ translation="MEMDKKVLNLRNSSLPLDLIRLLSDKDKDKIHEAWNYVIS
NVRDKDFLELLSLDGTGRYRANWSVPEFVERGILTEEVIRKEHFLMLKDSNKV
VRALSWYVTLKPLLEMNVSGLGVLSYSPFCELINSEPHVEVGVNQEFKITKPI"
/ 6684. 7412
/ gene="SSO1954"
/ 6684. .7412

```

```

/ gene="SSO1954"
/ note="Cell Envelope, Surface polysaccharides and
lipopolysaccharides"
/ codon_start=1
/ transl_table=11
/ product="Polichyl-phosphate mannose synthase related
protein"
/ protein_id="AAK42146.1"
/ db_xref="GI:13815228"
/ translation="WISIIIPAFNEERIGETLEKISTLPLENAEIVVVDGHDNTPEV
VAKFPVKLIISESRLGKGCALKRGINESNFORVLLDADPITEEELDKILSTADIV
IPKRIKIVGPKRKFHLHGFIITIKLFPFLAKFSDPGQKLVNRRKYLIDILBELII
NDPLFVNLIVGFKRHHYKVEVEINYIHDESDSKISKIKLIVILMFLSLIKURVYI
SPFRKILYTKYLKAQDYILRKLK"
/ complement (7395. .7934)
/ gene="SSO1956"
/ complement (7395. .7934)
/ gene="SSO1956"
/ note="Multicopy. Similar to SSO1359"
/ codon_start=1
/ transl_table=11
/ product="Hypothetical protein"
/ protein_id="AAK42147.1"
/ db_xref="GI:13815229"
/ translation="MIVTKRLLITISLLNVALLPGYVMYVSAGMVKDKSLPS
LYDEFLGKELEGIINILFAFRFVSVSLYIYALKDKDVINNILLITWISYLYL
FDPLLFYLLFNYYGVYVTPTKYPLFIIGSQNMTVFYKNNMTIIVESPTTYWIALF
AGTFNLISRIIISRLSKLS"
/ complement (7931. .8341)
/ gene="SSO1957"
/ complement (7931. .8341)
/ gene="SSO1957"
/ note="Multicopy. Similar to SSO1360"
/ codon_start=1
/ transl_table=11
/ product="Hypothetical protein"
/ protein_id="AAK42148.1"
/ db_xref="GI:13815230"
/ translation="MDLGIKILGVISLVIVQFAMNNFNGLLQTLNQINSTILL
FTSQPKIGLIVSNQILITFNLTNFPITLNVGTGDTYLSKPTINEVKNISVVV
TNLEMPERAIHDSNYLTIYMKMNTTFSQSEMI"
/ complement (8365. .9489)
/ gene="SSO1958"
/ complement (8365. .9489)
/ gene="SSO1958"
/ note="Transport, Cell membrane transport"
/ codon_start=1
/ transl_table=11
/ product="Transport membrane protein (permease)"
/ protein_id="AAK42149.1"
/ db_xref="GI:13815231"
/ translation="MLRNIIYSWTGTQLQLFLRLSGVYISLPALVHLSIQIGLVA
TVYFYGVSSTIPWGLIIDRGPSSAVYASILLGNMLLFLTSLTVAILLVYLIE
GLITAIIPFSAMKIVAVSYNSKRTFFVALVESAGPTIIITLGLISALLHWRFLYI
LIMASFGLLIAIFSHENDVNRTEIKESFKIILDRKTIATILRLGELSWTGTFTW
IPPMVLVNRNISPSTLSALFLFLGVGQVLISVERLVERFGDRTVILINLIGFILT
PSIIIPSNIDILPEAFLLIGTFSFVRPPTDSLIMRIAGQSAGTSIGYANAVSIQITM
IAPSFVGLTYLITHSFSISMLALDVGCIISIIISLLSKHL"
/ 9586. .10062
/ gene="SSO1959"
/ 9586. .10062
/ gene="SSO1959"
/ codon_start=1
/ transl_table=11
/ product="Hypothetical protein"
/ protein_id="AAK42150.1"
/ db_xref="GI:13815232"
/ translation="MPRELTPELQQLLLPGYRGYKAKDLIRQDDFLVRRVIDRLEQ

```

Query Match 4.9%; Score 47.2; DB 1; Length 11595;  
 Best Local Similarity 52.0%; Pred. No. 0.068;  
 Matches 106; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 589 GAAGTAATGATATGATTTACTATGACGGATTACACCGGCTGCTCCAAAGTTAAGGAG 648  
|||||  
Db 2633 GAGATAATGGTCGTGTTATTAGTCAGCAATTTACAGCGCTGCTGGGCAACAGTTGGAGAA 2574  
|||||  
QY 549 ATTCTAGTCCCAATAATAGTTAAACCGAACACAGTAATGCTACATTTGAAGTATGAAG 708  
|||||  
Db 2573 GTAGTAATCCCAATATATATTAATCACACTCTAGTCAACGCCACTTTCCCAAGTGTGGAAG 2514  
|||||  
QY 709 GCAACACATTTGGTGGGAGTATGTTGCATTTAGATAAAGACCCCAATCAAGAGGGGAACA 768  
|||||  
Db 2513 ATGAAGAACGTCCCATGGGAGGTGGGAGTACATAGACCATTTAGACCAGATGCGTGAAA 2454  
|||||  
QY 769 GTGACAATTCATACCGAGCATTT 792  
|||||  
Db 2453 GTCACAAATGGTTACGTCGCATAT 2430  
|||||

## RESULT 15

CEP17C11 37037 bp DNA linear INV 03-DEC-2003  
Caenorhabditis elegans cosmid F17C11, complete sequence.

DEFINITION Z72507  
VERSION Z72507.1 GI:11301657  
KEYWORDS HTG: 156G surface protein; bZIP; Collagen; Dihydrodipicolinate;  
Elongation factor 1-gamma; gamma; sepB.

SOURCE Caenorhabditis elegans  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

## REFERENCE 1

AUTHORS none.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
JOURNAL Science 282 (5396), 2012-2018 (1998)  
MEDLINE 99069613  
PUBMED 9851916  
REMARK The C.elegans Sequencing Consortium.  
REFERENCE 2 (bases 1 to 37037)  
AUTHORS McMurray,A.A.

JOURNAL Direct Submission  
TITLE Submitted (05-MAY-1996) Nematode Sequencing Project, Sanger  
Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwenematode.wustl.edu  
COMMENT Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.

Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
For a graphical representation of this sequence and its analysis  
see:- [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F17C11)  
name=F17C11

IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone F17C11.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.

The true left end of clone F17C11 is at 23758 in

sequence CEF58B4.

The true right end of clone F17C11 is at 8546 in

sequence CEF58B7.

The start of this sequence (1..104) overlaps with the end of

sequence CEF58B4.

The end of this sequence (36934..37037) overlaps with the start of  
sequence CEF58B7.

## FEATURES

Location/Qualifiers

1..37037

source

## ORIGIN

Query Match 4.8%; Score 46.4; DB 3; Length 37037;  
Best Local Similarity 55.6%; Pred. No. 0.12;  
Matches 89; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 35 TAACAATCCTTTTAGTACAGGCAATATATTTTGTAGAAAAGTATCATACCTCTGAGGACA 94  
30255 TCACGAATTTTCAACAACACAAATCTCTGAAAAAATGATTATTTATCAGTTTACA 30314  
QY 95 AGTCAACTTTCAAATACCTCATCTACACCAACCCCAACACACTTTTCCACTACCAAGGTTTC 154  
30315 CATTTACCTAAATTTACAATTCATCACTATTCCAAACAGCAATTTGATTATCCATGTTG 30374  
QY 155 TCAAGATTAGATACCTGATGACGGTGAGTGCCGAGGAGC 194  
30375 AGGAGGCAAGATACCTTCCACACACGGCGAAAGCGCAGTAAC 30414  
Db

Search completed: July 1, 2004, 23:39:57

Job time : 6094 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 20:10:44 ; Search time 22 Seconds  
(without alignments)  
748.577 Million cell updates/sec

Title: US-09-914-543-46  
Perfect score: 1701  
Sequence: 1 MSKKFVIVSILTLVQAI.....HLEWITNITPLDRPLIS 319

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: /cg2\_6/ptodata/2/iaa/5A COMB.pcp.\*
- 2: /cg2\_6/ptodata/2/iaa/5B COMB.pcp.\*
- 3: /cg2\_6/ptodata/2/iaa/6A COMB.pcp.\*
- 4: /cg2\_6/ptodata/2/iaa/6B COMB.pcp.\*
- 5: /cg2\_6/ptodata/2/iaa/6C COMB.pcp.\*
- 6: /cg2\_6/ptodata/2/iaa/backfiles.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	3	US-09-215-042-1
2	1701	100.0	319	4	US-09-134-078-64
3	486.5	28.6	288	3	US-08-985-280C-2
4	486.5	28.6	288	3	US-09-215-042-2
5	143.5	8.4	261	3	US-08-110-359A-2
6	128	7.5	260	3	US-09-216-295-23
7	128	7.5	260	4	US-09-632-570-23
8	128	7.5	260	4	US-09-632-575-53
9	119.5	7.0	553	1	US-08-651-572-2
10	119.5	7.0	553	3	US-09-066-544-2
11	119.5	7.0	553	3	US-08-981-086-2
12	119.5	7.0	553	4	US-09-430-669-2
13	114.5	6.7	381	3	US-09-216-295-22
14	114.5	6.7	381	4	US-09-632-570-22
15	114.5	6.7	381	4	US-09-632-575-52
16	114	6.7	233	1	US-08-032-848C-12
17	112.5	6.6	264	4	US-09-632-570-24
18	112.5	6.6	264	4	US-09-632-575-54
19	111	6.5	239	3	US-08-849-751-2
20	111	6.5	239	4	US-08-478-816-2
21	111	6.5	667	1	US-08-471-033-7
22	111	6.5	667	1	US-08-471-044-7
23	111	6.5	667	2	US-08-463-483A-7
24	111	6.5	667	2	US-08-471-046A-7
25	111	6.5	667	2	US-08-470-566B-7
26	111	6.5	667	2	US-08-469-334-7
27	111	6.5	667	3	US-09-300-529-7

28	108	6.3	263	3	US-09-216-295-24	Sequence 24, Appl
29	108	6.3	312	3	US-09-216-295-21	Sequence 21, Appl
30	108	6.3	371	3	US-08-104-308-1	Sequence 1, Appl
31	108	6.3	371	3	US-09-321-981-1	Sequence 1, Appl
32	108	6.3	371	4	US-09-739-861A-1	Sequence 1, Appl
33	108	6.3	371	4	US-09-795-583-1	Sequence 21, Appl
34	108	6.3	371	4	US-09-632-570-21	Sequence 21, Appl
35	108	6.3	371	4	US-09-632-575-51	Sequence 51, Appl
36	108	6.3	386	3	US-09-321-991-5	Sequence 5, Appl
37	108	6.3	386	4	US-09-739-861A-5	Sequence 5, Appl
38	108	6.3	386	4	US-09-795-583-5	Sequence 5, Appl
39	103	6.1	852	1	US-08-471-033-36	Sequence 36, Appl
40	103	6.1	852	2	US-08-471-044-36	Sequence 36, Appl
41	103	6.1	852	2	US-08-463-483A-36	Sequence 36, Appl
42	103	6.1	852	2	US-08-471-046A-36	Sequence 36, Appl
43	103	6.1	852	2	US-08-470-566B-36	Sequence 36, Appl
44	103	6.1	852	2	US-08-469-334-36	Sequence 36, Appl
45	103	6.1	852	3	US-09-300-529-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1  
US-09-215-042-1  
; Sequence 1, Application US/09215042  
; Patent No. 6426698  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jiyin  
; APPLICANT: Condon, Brian  
; TITLE OF INVENTION: Continuous Biopolishing of  
; TITLE OF INVENTION: Cellulose-Containing Fabrics With Thermophilic Cellulases  
; FILE REFERENCE: 5464.200-US  
; CURRENT APPLICATION NUMBER: US/09/215,042  
; CURRENT FILING DATE: 1998-12-17  
; EARLIER APPLICATION NUMBER: 60/068,274  
; EARLIER FILING DATE: 1997-12-19  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-215-042-1

Query Match	100.0%;	Score 1701;	DB 3;	Length 319;
Best Local Similarity	100.0%;	Pred. No. 3.7e-167;	Mismatches 0;	Indels 0; Gaps 0;
Matches 319;	Conservative 0;			
QY	1	MSKKFVIVSILTLVQAIYFVKYHTS	EDKSTNTSSTPQTTLSTTKVKIRYPDDG	60
Db	1	MSKKFVIVSILTLVQAIYFVKYHTS	EDKSTNTSSTPQTTLSTTKVKIRYPDDG	60
QY	61	EWFGAPIDKGGNPEFYIEINLNLNATGA	FTYNTLSTGVLHYVQOLDNIVLRDSN	120
Db	61	EWFGAPIDKGGNPEFYIEINLNLNATGA	FTYNTLSTGVLHYVQOLDNIVLRDSN	120
QY	121	WHGYPEIFYGNKPNANVATDGPILPSKVN	LDLDFYLTISYKLEPKNGLPINFAISW	180
Db	121	WHGYPEIFYGNKPNANVATDGPILPSKVN	LDLDFYLTISYKLEPKNGLPINFAISW	180
QY	181	LTRAWRTTGINSDEQVMWIIYYDGLQFAG	SKVKEIIVVPIIVNGTVPVNATEVWKANIG	240
Db	181	LTRAWRTTGINSDEQVMWIIYYDGLQFAG	SKVKEIIVVPIIVNGTVPVNATEVWKANIG	240
QY	241	WEYVAFRIKPTKEGTVTIPYCAFI	SVAAINISLNNYLYLEDVEICTEGTPTTSAH	300
Db	241	WEYVAFRIKPTKEGTVTIPYCAFI	SVAAINISLNNYLYLEDVEICTEGTPTTSAH	300
QY	301	LEWMIITNITLPLDRPLIS	319	
Db	301	LEWMIITNITLPLDRPLIS	319	

```
RESULT 2
US-09-134-078-64
; Sequence 64, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: Fast-SEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 319 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-134-078-64

Query Match 100.0%; Score 1701; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.7e-167;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSLTLLVQAIFYVEKHHTSBDKSTNTSSTPPQTTSTTKVLKIRYPDDG 60
Db 1 MSKKFVIVSLTLLVQAIFYVEKHHTSBDKSTNTSSTPPQTTSTTKVLKIRYPDDG 60

QY 61 EWPAPDKDGDGNEPFIENLWNLNATGFAEMTYNLTSGLVHYVQQLDNIVLRDSN 120
Db 61 EWPAPDKDGDGNEPFIENLWNLNATGFAEMTYNLTSGLVHYVQQLDNIVLRDSN 120

QY 121 WWHGYPEIFYGNKPWNANYATDGPILPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
Db 121 WWHGYPEIFYGNKPWNANYATDGPILPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

QY 181 LTREAWTTGNSDEQEWMIWYDGLQAGSKVKEIWPVPIVNGTPVNAFTFVKANIG 240
Db 181 LTREAWTTGNSDEQEWMIWYDGLQAGSKVKEIWPVPIVNGTPVNAFTFVKANIG 240

QY 241 WEYVAFRIKPIKEGTVTIPYGAFISVAANISLNPNTLYLEDEVEIGTFGTPSTSAH 300
Db 241 WEYVAFRIKPIKEGTVTIPYGAFISVAANISLNPNTLYLEDEVEIGTFGTPSTSAH 300

QY 301 LEWWTITNLTPLDRPLIS 319
Db 301 LEWWTITNLTPLDRPLIS 319

RESULT 3
US-08-995-280C-2
; Sequence 2, Application US/08995280C
; Patent No. 6043075
; GENERAL INFORMATION:
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Schulein, Martin
; APPLICANT: No. 6043075revang, Iben Angelica
; TITLE OF INVENTION: A No. 6043075el Endoglucanase
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6043075o No. 6043075disk of No. 6043075th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,280C
; FILING DATE: 19-Dec-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol A.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4906.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-995-280C-2

Query Match 28.6%; Score 486.5; DB 3; Length 288;
Best Local Similarity 32.2%; Pred. No. 7.6e-42;
Matches 104; Conservative 58; Mismatches 106; Indels 55; Gaps 5;

QY 4 KKFVIVSLTLLVQAIFYVEKHHTSBDKSTNTSSTPPQTTSTTKVLKI 54
Db 2 KKSLLSLILILLITLSFSQTPKYKADAFILKAPSSGDVTTKMLPLT----- 47

QY 55 RYDDGDFGAPDKDGDGNEPFIENLWNLNATGFAEMTYNLTSGLVHYVQQLDNIV 114
Db 48 -----LELFWNIANVEGNTWMAFYKEEDTVEYYADIKNIV 83

QY 115 LRDRNNWVHGPEIFYGNKPWNANYATDGPILPSKVSNTLDFYLTISYKLEPKNGLPIN 174
Db 84 LKDKNSWVHGPEIFYGYKPKWAGHNSIEKLALPKVSEFPDVLNFKLNIWYKELPIN 143

QY 175 FAIBSLTREATRTTGINSDEQEWMIWYDGLQAGSKVKEIWPVPIVNGTPVNAFTFV 234
Db 144 FAMEWTIKPEYQKT-VTSGDIEMWMLYANLSPAGKVGKIPILNNGNQDKIWEV 202

QY 235 WKANIGWYVAFRIKPIKEGTVTIPYGAFIS-----VAANISL--PNTELYLEDEVEI 287
Db 203 YLSPMSWDYVAYKSKENILQGVKIPINEFLKRLTILANPNPSRITPEKFDQMTVTWVEI 262

QY 288 GTEGCTPSTTSAHLEWWTITNLT 310
Db 263 GTEFGDPYTTEAKFGWTFNSFDI 285
```



```
RESULT 4
US-09-215-042-2
; Sequence 2, Application US/09215042
; Patent No. 6126698
; GENERAL INFORMATION:
; APPLICANT: Liu, Jiyin
; APPLICANT: Condon, Brian
; TITLE OF INVENTION: Continuous Biopolishing of
; TITLE OF INVENTION: Cellulose-Containing Fabrics With Thermophilic Cellulases
; FILE REFERENCE: S464.200-US
; CURRENT APPLICATION NUMBER: US/09/215.042
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/068,274
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Dictyoglomus sp.
US-09-215-042-2

Query Match      28.6%; Score 486.5; DB 3; Length 288;
Best Local Similarity 32.2%; Pred. No. 7.6e-42;
Matches 104; Conservative 58; Mismatches 106; Indels 55; Gaps 5;

QY 4 KKFVIVGILITLLVQAIV-----FVEKYHTSEDKSTSTSTSTPPQTLSTTKYLKI 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 KKSLLSLLILLTLTSLFSQTPPKYKDAFILKAPSSGDVTTKNLPLT----- 47

QY 55 RYPDGGWPGAPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIV 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 48 -----LELFWNIANYEGNTWMAFYKEEDTVEYYADIKNIY 83

QY 115 LDRSSNVHGYPEIFYGNKPNANYATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPIN 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 LKXNSVHGYPEYVYGYKPGWAGHNSIEKIALPKKSEFPDVLNFKYNIWEKULPIN 143

QY 175 FAIESWLTREAWRTTGINSDEQEVMIWIYDGLQPGAGSKYKEIVVPIIVNGTTPVNATFEV 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 144 FAMEWTITKEPYQKT-VTSGDIEMWVLYANRLSPAGRKVGKIPILNGNQKDIWEV 202

QY 235 WKANIGWEYVAFRIKTIKEGTVTIPYGATIS-----VAANISSL--PNYTELYLEDVEI 287
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 203 YLSPMSWDYVAYKSENILQGVKIPINEFUKHLRTILANPNPSKITPEKFDQMYVTWEI 262

QY 288 GTEFGTPTSTSAHLEWITNITL 310
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 263 GTEFGDPVTTEAKGWTFSNFDI 285

RESULT 5
US-09-110-959A-2
; Sequence 2, Application US/09110959A
; Patent No. 6268197
; GENERAL INFORMATION:
; APPLICANT: Schulein, Martin
; APPLICANT: Outtrup, Helie
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Bjornvad, Mads Eskelund
; TITLE OF INVENTION: Alkaline Xyloglucanase
; FILE REFERENCE: 5206.200-US
; CURRENT APPLICATION NUMBER: US/09/110.959A
; CURRENT FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 0822/97
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: 1213/97
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/054,039
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 60/063,694
; PRIOR FILING DATE: 1997-10-28
```

```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Bacillus licheniformis ATCC 14580
US-09-110-959A-2

Query Match      8.4%; Score 143.5; DB 3; Length 261;
Best Local Similarity 23.9%; Pred. No. 1.7e-06;
Matches 63; Conservative 37; Mismatches 97; Indels 67; Gaps 15;

QY 66 PIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSRNW---- 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 PSDKLYFKNKYYIFNNVWGADQVSGWQTIYHNSDSMDGWVW-----NWPSNT 83

QY 122 --VHGYPEIFYGNKPNANYATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIES 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 STVKAYPSIVSG-WHWTGTTAGS--GFTTRUSDQKNINTRKVSIS-ANG-TYNAAYDI 138

QY 180 WL---TREAWRTTGINSDEQEVMIWIYDGLQPGAGSKYKEIVVPIIVNGTTPVNATFEVWK 236
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 WLHNTNKASWD---SAPTDEIMIMLNNTNAGPAGSYVETVSIG-----GHSWKVYK 186

QY 237 ANI-----GWEYVAFRIKTIKEGTVTIPYGAFISVAANISSLPNY-----TELY 281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 GYIDAGGKGWNVFSG-IRTANTQ-----SANLIRDFNTLYLADSKQWLSKTKY 234

QY 282 LEDVEIGTE-FGTPSTTSAHLEW 304
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 VSSVEFGTEVFG--GTGQINISNW 256

RESULT 6
US-09-216-295-23
; Sequence 23, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Rhodothermus marinus
US-09-216-295-23

Query Match      7.5%; Score 128; DB 3; Length 260;
Best Local Similarity 24.1%; Pred. No. 6.7e-05;
Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;

QY 59 DGEWPGAPIDKDGNGNPEFYIEI-----NLWNILNATGFAEMTYNLTSGVLHYV 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 DMLFPDGDNGKEPEPEPEPTVELCGRWDAWDVAGGRYRVNNVWGAEQAQCIIEVL---- 74

QY 108 QQLDNIVL-----RDRSNVWHGYPEIFYGNKPNANYATDGPILPSPKSVNLTDFYLTISY 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 -ETGNFTITRADHDNGNNVAAYPAYFVGCH-WAPARAIRDCARAGAVRRRAHELDVT--- 129

QY 164 KLEPKNGLP-----NFAIESWLTREAWRTTGINSDEQEVMIWIYD-GLQPGAGSKYKEI 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 -----PITGRNNAAYDIWFSPVTNSGNGY-SGGAELMIWLWNGGWMPEGGRVATV 180

QY 218 VVPIIVNGTTPVNATFEVWKANIGWEYVAFRIKTIKEGTVTIPYGAFISVAANISSLPNY 277
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 EL-----AGATWEVYVADWDVNIYVARRTTP-TTSVSELDLAKFIDDAVARGVI--R 229

QY 278 TELYLEDVEIGTE 290
```



Db 374 TTTTSTPPTTTTSTPTTTTQTPTTTTPTTTTTPSNNVPEIVNV-LPTSSQ 432  
QY 62 WPGAPIDKDGDN---PEFYIENLWNIL---NAT-----GFAEMT 96  
Db 433 YEGTSVEVCDGTQACSSVWGAPNLGVVKGIGNATMDPNVGMWEDVYKTAPODGTGSK 492  
QY 97 YNLTSGLVHYVQQLDNVLRDRSNVHGYPFIYGNKPMNANYATDGP-IPLPSKVSNL 154  
Db 493 MEIRNGVLK-VTNLWNINHPKNTM-AYPEVIYGAKPW-GNQPINAPNFVLPKVSQ 548

## RESULT 10

US-09-066-544-2  
; Sequence 2, Application US/09066544  
; Patent No. 6001984  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,544  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-09-066-544-2

Query Match 7.0%; Score 119.5; DB 3; Length 553;  
Best Local Similarity 25.1%; Pred. No. 0.0017;  
Matches 45; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

QY 33 STNTSSTPPTTLST-----KVLKIRYPDGE 61  
Db 374 TTTTSTPPTTTTSTPTTTTQTPTTTTPTTTTTPSNNVPEIVNV-LPTSSQ 432  
QY 62 WPGAPIDKDGDN---PEFYIENLWNIL---NAT-----GFAEMT 96  
Db 433 YEGTSVEVCDGTQACSSVWGAPNLGVVKGIGNATMDPNVGMWEDVYKTAPODGTGSK 492  
QY 97 YNLTSGLVHYVQQLDNVLRDRSNVHGYPFIYGNKPMNANYATDGP-IPLPSKVSNL 154  
Db 493 MEIRNGVLK-VTNLWNINHPKNTM-AYPEVIYGAKPW-GNQPINAPNFVLPKVSQ 548

## RESULT 11

US-08-951-086-2  
; Sequence 2, Application US/08951086  
; Patent No. 6074867  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/951,086  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/651,572  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-951-086-2

Query Match 7.0%; Score 119.5; DB 3; Length 553;  
Best Local Similarity 25.1%; Pred. No. 0.0017;  
Matches 45; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

QY 33 STNTSSTPPTTLST-----KVLKIRYPDGE 61  
Db 374 TTTTSTPPTTTTSTPTTTTQTPTTTTPTTTTTPSNNVPEIVNV-LPTSSQ 432  
QY 62 WPGAPIDKDGDN---PEFYIENLWNIL---NAT-----GFAEMT 96  
Db 433 YEGTSVEVCDGTQACSSVWGAPNLGVVKGIGNATMDPNVGMWEDVYKTAPODGTGSK 492  
QY 97 YNLTSGLVHYVQQLDNVLRDRSNVHGYPFIYGNKPMNANYATDGP-IPLPSKVSNL 154  
Db 493 MEIRNGVLK-VTNLWNINHPKNTM-AYPEVIYGAKPW-GNQPINAPNFVLPKVSQ 548

## RESULT 12

US-09-430-669-2  
; Sequence 2, Application US/09430669  
; Patent No. 6329187  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND



```

; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces lividans CelB
US-09-632-575-52

Query Match      6.7%; Score 114.5; DB 4; Length 381;
Best Local Similarity 27.1%; Pred. No. 0.003;
Matches 52; Conservative 21; Mismatches 80; Indels 39; Caps 10;

Qy 125 YPELFYGNKPNANYATDGPPLPSKVSNLTFYLTISYKLEPKNGLPINFAIESWLTRE 184
Db 97 YPSVFNGCHYTNCSFGTD----LPVRLDTVSAAPSSISYGF--VDGAVYNASYDIWLDPT 150

Qy 185 AWRRTGINSDEQEVMIWYYDG-LQPAGSKYKEIWPFIYNGTPV--NATFEVM----- 235
Db 151 A-RTDGVN--QTEIMWFNRVGPQIPIGSPV-----GTASVGGRTWEVMSGNGS 197

Qy 236 -----KANIGWEY-VAPRIKTIKEGTVTIPYGAFTSVAANISSLBNYTELYEDV 285
Db 198 NDVLSFVAPSAISGWSFDMDFVRATVARGLAENDW-YLTSVQAGPEPQNGAGLAVNSF 256

Qy 286 EIGTEFGTPTST 297
Db 257 SSTVETGTPGCT 268
```

Search completed: June 29, 2004, 20:14:10  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 10:56:46 ; Search time 603 Seconds  
(without alignments)  
6763.300 Million cell updates/sec

Title: US-09-914-543-45  
Perfect score: 960  
Sequence: 1 atgagcagaagaaagtcgt.....atagacctttatttcttaa 960

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002s:\*  
7: Geneseq2003as:\*  
8: Geneseq2003bs:\*  
9: Geneseq2003cs:\*  
10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	960	2 AAT94215	Aat94215 Pyrococcus
2	960	100.0	960	2 AAV47540	Aav47540 Pyrococcus
3	960	100.0	960	2 AAV36924	Aav36924 Pyrococcus
4	960	100.0	960	9 ADC26951	Adc26951 Pyrococcus
5	958.4	99.8	1134	6 ABA91905	Ab91905 Pyrococcus
6	958.4	99.8	1134	6 ABA00497	Ab00497 P. furios
7	536.8	55.9	978	7 ACC44578	Acc44578 pNOV4800
8	534	55.6	903	7 ACC44577	Acc44577 Pyrococcus
9	115.8	12.1	2529	2 AAT94193	Aat94193 Archaeabac
10	114.2	11.9	894	2 AAV40651	Aav40651 Endo-1,4-
11	114.2	11.9	894	2 AAV47544	Aav47544 pMB47A p
12	47.4	4.9	2000	7 ADA71938	Ada71938 Rice gene
13	47.2	4.9	999	2 AAX80187	Aax80187 Sulfolobu
14	44.2	4.6	999	2 AAX80188	Aax80188 Sulfolobu
15	43.6	4.5	53585	2 AAX20251	Aax20251 Borrelia
16	38.6	4.0	2150	6 ABA85746	Abk85746 DNA encod
17	37.8	3.9	728	2 AAV74922	Aav74922 Staphyloc
18	37.8	3.9	1341	7 ACF74977	Acf74977 Staphyloc
19	37.2	3.9	2338	6 ABA63651	Abk63651 Rat sequ
20	37.2	3.9	2338	6 ABA58086	Ab58086 Toxicity-
21	37.2	3.9	2338	9 ABA52565	Ab52565 Primary r
22	36.6	3.8	469	5 ABA52725	Abv52725 Human pro
23	36.6	3.8	110000	6 ABA90521_21	Continuation (22 o

24	36.4	3.8	741	2 AAX98933	Aax98933 Human val
c 25	36.4	3.8	25580	2 AAX12956	Aax12956 Enterococ
c 26	36.4	3.8	25580	6 ABA98751	Ab98751 Enterococ
27	36.2	3.8	2000	7 ADA71938	Ada71938 Rice gene
28	36	3.8	936	4 AAF75659	Aaf75659 Human ztr
29	36	3.8	936	8 ACD25921	Ac25921 Degenerat
30	36	3.8	936	9 ADC17391	Adc17391 Human ear
c 31	36	3.8	12393	6 ABL33263	Ab133263 Human imm
c 32	35.6	3.7	497	6 ABA45944	Ab45944 CDNA enco
33	35.6	3.7	506	6 ABA87188	Abv87188 Human col
34	35.6	3.7	1083	7 ACA29166	Aca29166 Prokaryot
35	35.4	3.7	2568	7 AAX84329	Aax84329 Stealth v
c 36	35.4	3.7	26747	6 AA167784	Aa167784 Nucleotid
37	35.2	3.7	496	5 AAH81653	Aah81653 Human dif
38	35.2	3.7	1492	5 AAS86756	Aas86756 DNA encod
39	35.2	3.7	1548	4 AA159507	Aa159507 Human pol
40	35	3.6	14041	4 AAH48024	Aah48024 Internal
c 41	34.8	3.6	3146	4 AA158397	Aa158397 Human pol
c 42	34.8	3.6	3146	8 ABA848367	Ab48367 Novel hum
c 43	34.8	3.6	3150	6 ABL65914	Ab165914 Lung canc
c 44	34.8	3.6	3150	6 ABL69587	Ab169587 Prostate
c 45	34.8	3.6	3150	6 ABL64453	Ab164453 Stomach c

## ALIGNMENTS

### RESULT 1

AAT94215	AAT94215 standard; DNA; 960 BP.
ID	
XX	AC AAT94215;
XX	AC AAT94215;
DT	17-OCT-2003 (revised)
DT	21-MAY-1998 (first entry)
XX	XX Pyrococcus furiosus endoglucanase DNA.
DE	XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
XX	KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW	thermostable enzyme; thermophilic; glycosidase; ss.
XX	XX Pyrococcus furiosus; (clone 7EG1).
OS	XX Pyrococcus furiosus; (clone 7EG1).
XX	XX WO9744361-A1.
PN	XX 27-NOV-1997.
XX	XX 27-NOV-1997.
XX	XX 22-MAY-1997; 97WO-US008793.
XX	XX 22-MAY-1996; 96US-00651572.
PR	(RECO-) RECOMBINANT BIOTOCATALYSIS INC.
PA	Lam DE, Mathur EJ;
XX	XX Lam DE, Mathur EJ;
PI	XX WPI; 1998-018435/02.
XX	XX P-PSDB; AAW35007.
DR	XX Endo:glucanase(s), preferably form archaeal bacterium, AEPII la - useful
XX	to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
PT	bonds in cellulose.
PT	XX Claim 3; Fig 1W; 164pp; English.
XX	XX This DNA sequence from Pyrococcus furiosus (clone 7EG1) encodes an
XX	endoglucanase (see AAW35007) that is able to degrade
CC	carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic bonds in
CC	cellulose, and which shows homology to the thermostable endoglucanase
CC	(see AAW34985) of archaeobacterium hydrothermal vent isolate AEP11a. The
CC	DNA can be used in the recombinant production of the endoglucanase and as
CC	a probe to identify similar sequences. 24 Endoglucanase polynucleotides
CC	(see AAT94193-216) are claimed. These can be incorporated into plasmid or





QY 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAAACTTCAAAATACCTCATCTACA 120  
 Db |||||  
 QY 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAAACTTCAAAATACCTCATCTACA 120  
 Db |||||  
 QY 121 CCACCCCAACAAACATTTTCACTACCAAGTTTCAAGATTAGATACCTGTAGCGGT 180  
 Db |||||  
 QY 121 CCACCCCAACAAACATTTTCACTACCAAGTTTCAAGATTAGATACCTGTAGCGGT 180  
 Db |||||  
 QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATTTTACATTGAA 240  
 Db |||||  
 QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATTTTACATTGAA 240  
 Db |||||  
 QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300  
 Db |||||  
 QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300  
 Db |||||  
 QY 301 AGCGGCTCTTCACTACGTCACAACTTCCACACATTTGCTGAGGATAGAAATTAAT 360  
 Db |||||  
 QY 301 AGCGGCTCTTCACTACGTCACAACTTCCACACATTTGCTGAGGATAGAAATTAAT 360  
 Db |||||  
 QY 361 TGGGTGATGATACCCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420  
 Db |||||  
 QY 361 TGGGTGATGATACCCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420  
 Db |||||  
 QY 421 ACTGATGCCCCAATACCAATTTACCCAGTAAAGTTTCAAACTTACAGACTTCTATTAACA 480  
 Db |||||  
 QY 421 ACTGATGCCCCAATACCAATTTACCCAGTAAAGTTTCAAACTTACAGACTTCTATTAACA 480  
 Db |||||  
 QY 481 ATCTCTTATAAATTTGACCCCAAGAACGGCTGCAATTAATCTTCCAAATAGAACTCTGG 540  
 Db |||||  
 QY 481 ATCTCTTATAAATTTGACCCCAAGAACGGCTGCAATTAATCTTCCAAATAGAACTCTGG 540  
 Db |||||  
 QY 541 TTAACGAGAGAAGCTTGGAGAACCAACAGGAATTAACGCGATGAGCAAGTAATGATA 600  
 Db |||||  
 QY 541 TTAACGAGAGAAGCTTGGAGAACCAACAGGAATTAACGCGATGAGCAAGTAATGATA 600  
 Db |||||  
 QY 601 TGGATTTACTATGAGGATTAACAACGGCTGCTCCAAAGTTTAAAGGAGTTGTAGTCCA 660  
 Db |||||  
 QY 601 TGGATTTACTATGAGGATTAACAACGGCTGCTCCAAAGTTTAAAGGAGTTGTAGTCCA 660  
 Db |||||  
 QY 661 ATAATAGTAAACGGAACACCACTAAATGCTACATTTGAAGTATGGAAGGCAAACTTGGT 720  
 Db |||||  
 QY 661 ATAATAGTAAACGGAACACCACTAAATGCTACATTTGAAGTATGGAAGGCAAACTTGGT 720  
 Db |||||  
 QY 721 TGGGATGATGTTGATTTAGATTAAGACCCCAATCAAGAGGAAACAGTCACAATTTCCA 780  
 Db |||||  
 QY 721 TGGGATGATGTTGATTTAGATTAAGACCCCAATCAAGAGGAAACAGTCACAATTTCCA 780  
 Db |||||  
 QY 781 TACGAGCATTTATAAGTGTTCGAGCCAAACATTTCAAGCTTTACCAAAATTTACAGAACTT 840  
 Db |||||  
 QY 781 TACGAGCATTTATAAGTGTTCGAGCCAAACATTTCAAGCTTTACCAAAATTTACAGAACTT 840  
 Db |||||  
 QY 841 TACTTAGAGGACGTGGAGATTTGGAACCTGAGTTTGGAAACGCGCAAGCACTACCTCGCCAC 900  
 Db |||||  
 QY 841 TACTTAGAGGACGTGGAGATTTGGAACCTGAGTTTGGAAACGCGCAAGCACTACCTCGCCAC 900  
 Db |||||  
 QY 901 CTAGATGGTGGATCACAAACATTAACACTAACTCTCTAGATAGACCTCTTATTTCTAA 960  
 Db |||||  
 QY 901 CTAGATGGTGGATCACAAACATTAACACTAACTCTCTAGATAGACCTCTTATTTCTAA 960  
 Db |||||

## RESULT 3

AAV36924

ID AAV36924 standard; DNA; 960 BP.

XX AC AAV36924;

XX XX

DT 17-OCT-2003 (revised)

DT 21-DEC-1998 (first entry)

XX XX

DE Pyrococcus VC1-7EG1 glycosidase gene coding region.

XX Glycosidase; VC1-7EG1; thermostable enzyme; oligosaccharide; glucose;

XX Glycosidase; VC1-7EG1; thermostable enzyme; oligosaccharide; glucose;

XX sugar; baking; textile; detergent; beta-galactosidase; ss.  
 OS Pyrococcus furiosus; strain VC1-7EG1.

PN MO9824799-A1.

XX 11-JUN-1998.

XX 08-DEC-1997; 97WO-US022623.

XX 06-DEC-1996; 96US-0056916P.

PR 10-OCT-1997; 97US-00949026.

XX XX

PA (DIVE-) DIVERSA CORP.

XX XX

PI Bylina EJ, Swanson RV, Mathur EJ, Lam DE;

XX WPI; 1998-352407/31.

DR P-PSDB; AAW49875.

XX XX

PS Claim 1; Fig 18a-b; 92pp; English.

XX XX

CC This isolated polynucleotide comprises a coding region for glycosidase

CC VC1-7EG1 (see AAW49875) from a clone (7EG1) of Pyrococcus furiosus VC1,

CC which grows optimally at 100 degC. The invention provides 18

CC polynucleotides (see AAV36907-24) coding for thermostable glycosidases

CC (see AAW49558-75) having glucosidase, alpha-galactosidase, beta-

CC galactosidase, beta-mannosidase, beta-mannanase, endoglucanase or

CC pullulanase activity. Vectors and host cells are also claimed. A method

CC is provided for producing the enzymes by recombinant techniques. A

CC claimed method for generating glucose from soluble cell oligosaccharides

CC comprises contacting a sample (selected from dairy products, fruit juice,

CC detergent, textile, guar gum, animal feed, plant biomass or waste

CC product) containing oligosaccharides (selected from maltose, cellobiose,

CC lactose, sucrose, raffinose, stachyose, verbascose, cellobiose, starch,

CC amylose, glycogen, disaccharides, polysaccharides and pullulan) with one

CC of the claimed glycosidases such that glucose is produced. (Updated on 17

CC -OCT-2003 to standardise OS field)

XX XX

SQ Sequence 960 BP; 329 A; 207 C; 186 G; 238 T; 0 U; 0 Other;

Query Match 100.0%; Score 960; DB 2; Length 960;

Best Local Similarity 100.0%; Pred. No. 1.4e-276;

Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAGTTGCTCATCTATCTTAAACAATCCTTTTAGTACAGCAATA 60

Db |||||

QY 1 ATGAGCAAGAAAAGTTGCTCATCTATCTTAAACAATCCTTTTAGTACAGCAATA 60

Db |||||

QY 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAAACTTCAAAATACCTCATCTACA 120

Db |||||

QY 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAAACTTCAAAATACCTCATCTACA 120

Db |||||

QY 121 CCACCCCAACAAACATTTTCCACTACCAAGTTTCTCAAGATTAGATACCTGTAGCGGT 180

Db |||||

QY 121 CCACCCCAACAAACATTTTCCACTACCAAGTTTCTCAAGATTAGATACCTGTAGCGGT 180

Db |||||

QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATTTTACATTGAA 240

Db |||||

QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATTTTACATTGAA 240

Db |||||

QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300

Db |||||

QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300

Db |||||

QY 301 AGCGGCTCTTCACTACGTCACAACTTGGCAACATTTGCTTCCAGGATAGAAATTAAT 360

Db |||||

QY 301 AGCGGCTCTTCACTACGTCACAACTTGGCAACATTTGCTTCCAGGATAGAAATTAAT 360

Db |||||

QY 361 TGGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATGGAATGCAAACTACGCA 420  
 Db 361 TGGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATGGAATGCAAACTACGCA 420  
 QY 421 ACTGATGCCCAATACCAATACCCAGTAAAGTTTCAACCTTAACAGACTTCTATCTAACA 480  
 Db 421 ACTGATGCCCAATACCAATACCCAGTAAAGTTTCAACCTTAACAGACTTCTATCTAACA 480  
 QY 481 ATCTCTATAAATTTAGCCCAAGAACGGCTGCTCAATTAATCTTGCATAGAAATCTGG 540  
 Db 481 ATCTCTATAAATTTAGCCCAAGAACGGCTGCTCAATTAATCTTGCATAGAAATCTGG 540  
 QY 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAAGAGTAATGATA 600  
 Db 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAAGAGTAATGATA 600  
 QY 601 TGGATTTACTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660  
 Db 601 TGGATTTACTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660  
 QY 661 ATAATAGTAAACGGAACCCAGTAATCTTCAATTTGAAGTATGGAAGGCAACATTTGGT 720  
 Db 661 ATAATAGTAAACGGAACCCAGTAATCTTCAATTTGAAGTATGGAAGGCAACATTTGGT 720  
 QY 721 TGGGAGTATGTTGCAATTAAGATAAGACCCCAATCAAGAGGGAACAGTGACAAATCCA 780  
 Db 721 TGGGAGTATGTTGCAATTAAGATAAGACCCCAATCAAGAGGGAACAGTGACAAATCCA 780  
 QY 781 TACGAGAGCTTTATAGTGTTCAGCCCAACATTTCAAGCTTACCAATTAACAGAACTT 840  
 Db 781 TACGAGAGCTTTATAGTGTTCAGCCCAACATTTCAAGCTTACCAATTAACAGAACTT 840  
 QY 841 TACTTAGAGGAGCTGGAGATTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 900  
 Db 841 TACTTAGAGGAGCTGGAGATTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 900  
 QY 901 CTAGAGTGGTGATCACAACATACAACTAATCTCTCTAGATAGACCTCTTATTTCTTAA 960  
 Db 901 CTAGAGTGGTGATCACAACATACAACTAATCTCTCTAGATAGACCTCTTATTTCTTAA 960

RESULT 4  
 ADC26951  
 ID ADC26951 standard; DNA; 960 BP.  
 XX AC ADC26951;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE Pyrococcus furiosus glycosidase encoding DNA SEQ ID NO:60.  
 XX KW enzyme delivery matrix; thermostable enzyme; glycosidase; hydrolysis;  
 XX KW lactose; galactose; glucose; guar gum; animal feed; hydraulic fracturing;  
 XX KW oil recovery; gas recovery; biodegradable; enzyme; gene; ds.  
 XX OS Pyrococcus furiosus.  
 XX FH Location/Qualifiers  
 XX CDS 1..960  
 XX FT /\*tag= a  
 XX FT /product= "glycosidase"  
 XX PN WO2003072717-A2.  
 XX PD 04-SEP-2003.  
 XX PF 21-FEB-2003; 2003WO-US005189.  
 XX PR 21-FEB-2002; 2002US-00081475.  
 XX PA (DIVE-) DIVERSA CORP.  
 XX PI Miller CA;

XX DR WPI; 2003-748224/70.  
 XX DR P-PSDB; ADC26955.  
 XX PT Enzyme delivery matrix used as adjuvant for animal feeds comprises  
 PT discrete particles each comprising grain germ that is spent of oil and  
 PT thermostable enzyme.  
 XX PS Example 1; SEQ ID NO 60; 103pp; English.  
 XX CC The present invention describes an enzyme delivery matrix (1) comprising  
 CC discrete particles each comprising grain germ that is spent of oil and a  
 CC thermostable enzyme, where the particle readily disperses the glycosidase  
 CC enzyme into aqueous media. Also described: (1) a method for preparing an  
 CC enzyme delivery matrix; (2) isolated nucleic acid molecules encoding the  
 CC enzymes; (3) utilising enzymes for hydrolysing guar gum; (5) nucleic acid  
 CC glucose; (4) utilising enzymes or polynucleotides to generate probes.  
 CC (1) can be used as an adjuvant for animal feeds, and for hydraulic  
 CC fracturing in oil and gas recovery. (1) is biodegradable. The present  
 CC sequence encodes a glycosidase enzyme, which is used in the  
 CC exemplification of the present invention.  
 XX SQ Sequence 960 BP; 329 A; 207 C; 186 G; 238 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 960; DB 9; Length 960;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-276;  
 Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGAGCAAGAAAAGTTGCTCATCTGTAATCTATCTTAACTCTTTTGTAGTACAGCAATA 60  
 Db 1 ATGAGCAAGAAAAGTTGCTCATCTGTAATCTATCTTAACTCTTTTGTAGTACAGCAATA 60  
 QY 61 TATTTTGTAGAAAAGTATCATCTCTGAGGACAAAGTCAACTTCAATACCTCTCATCTACA 120  
 Db 61 TATTTTGTAGAAAAGTATCATCTCTGAGGACAAAGTCAACTTCAATACCTCTCATCTACA 120  
 QY 121 CCACCCCAACAACTTTTCCACTACCAAGTTTCTCAAGATTAGATACCTGATACCGT 180  
 Db 121 CCACCCCAACAACTTTTCCACTACCAAGTTTCTCAAGATTAGATACCTGATACCGT 180  
 QY 181 GAGTGCCAGGAGCTCTCTATGATAGGATGGTATGGGACCCAGAAATTTCTACATTGAA 240  
 Db 181 GAGTGCCAGGAGCTCTCTATGATAGGATGGTATGGGACCCAGAAATTTCTACATTGAA 240  
 QY 241 ATAACCTTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGATACCTTAACC 300  
 Db 241 ATAACCTTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGATACCTTAACC 300  
 QY 301 AGCGGCTCTTCACTACGTCACCAACATTTGACCAATTTCTTGGGGATAGAGTAAT 360  
 Db 301 AGCGGCTCTTCACTACGTCACCAACATTTGACCAATTTCTTGGGGATAGAGTAAT 360  
 QY 361 TGGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATGGAATGCAAACTACGCA 420  
 Db 361 TGGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATGGAATGCAAACTACGCA 420  
 QY 421 ACTGATGGCCCAATACCATTTACCAAGTTTCAACCTTAACAGACTTCTATCTAACA 480  
 Db 421 ACTGATGGCCCAATACCATTTACCAAGTTTCAACCTTAACAGACTTCTATCTAACA 480  
 QY 481 ATCTCTATAAATTTAGCCCAAGAACGGCTGCTCAATTAATCTTGCATAGAAATCTGG 540  
 Db 481 ATCTCTATAAATTTAGCCCAAGAACGGCTGCTCAATTAATCTTGCATAGAAATCTGG 540  
 QY 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAAGAGTAATGATA 600  
 Db 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAAGAGTAATGATA 600  
 QY 601 TGGATTTACTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660  
 Db 601 TGGATTTACTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660

QY 661 ATAATAGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGAAGGCAAAACATTGGT 720  
 Db 661 ATAATAGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGAAGGCAAAACATTGGT 720  
 QY 721 TGGGAGTATGTCATTTAGATAAAGACCCCAATCAAAAGAGGGAACAGTGACAATTTCCA 780  
 Db 721 TGGGAGTATGTCATTTAGATAAAGACCCCAATCAAAAGAGGGAACAGTGACAATTTCCA 780  
 QY 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTTACACAGAACTT 840  
 Db 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTTACACAGAACTT 840  
 QY 841 TACTTAGAGGAGCTGGAGATTGGAACCTGAGTTTGAAGCCCAAGCACTACCTCCGCCAC 900  
 Db 841 TACTTAGAGGAGCTGGAGATTGGAACCTGAGTTTGAAGCCCAAGCACTACCTCCGCCAC 900  
 QY 901 CTAGAGTGGTGATCACAAACATACACTAACTCTCTAGATAGACCTCTATTTCCTAA 960  
 Db 901 CTAGAGTGGTGATCACAAACATACACTAACTCTCTAGATAGACCTCTATTTCCTAA 960

RESULT 5

ABA91905  
 ID ABA91905 standard; DNA; 1134 BP.

XX ABA91905;

AC ABA91905;

DT 15-MAY-2002 (first entry)

DE Pyrococcus furiosus thermostable cellulase egIA gene.

XX Cellulase; EgIA; thermostable; enzyme; gene; ds.

XX Pyrococcus furiosus.

FT Key Location/Qualifiers

FT CDS 82..1041

FT /\*tag= a

FT /product= "Cellulase\_EgIA"

FT sig\_peptide 82..138

FT /\*tag= b

FT mat\_peptide 139..1038

FT /\*tag= c

PN WO200196382-A2.

XX 20-DEC-2001.

XX 15-JUN-2001; 2001WO-ISO000012.

XX 15-JUN-2000; 2000US-00594884.

XX (PROK-) PROKARIA EHF.

XX Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;

XX WPI; 2002-226850/28.

XX P-PSDB; AAM50997.

XX Novel thermostable variant cellulase which is truncated such that amino  
 PT terminal hydrophobic region and linker group of corresponding full length  
 PT enzyme is deleted, has improved catalytic properties and/or stability.

XX Disclosure; Page 45-47; 51pp; English.

XX The present sequence is that of the *Pyrococcus furiosus* egIA gene  
 CC encoding thermostable cellulase EgIA (see AAM50997), a family 12 glycosyl  
 CC hydrolase. The invention provides polypeptides having thermostable  
 CC cellulase activity. These are variants of a glycosyl hydrolase of family  
 CC 12 and are truncated such that 1 or more of the amino acid residues  
 CC corresponding to position 1 to about 40 are deleted. The polypeptides are  
 CC particularly derived from thermophilic *Rhodothermus* and *Pyrococcus* spp.  
 CC For EgIA, amino acids from the N-terminal hydrophobic region and/or the

CC linking moiety are deleted. Claimed polypeptides (see AAM50976-996) are  
 CC also derived from *Rhodothermus marinus* Celli2A cellulase. Isolated nucleic  
 CC acids encoding these thermostable cellulase polypeptides, and host cells,  
 CC are also claimed. The thermostable variant cellulases are useful e.g. in  
 CC the wood and paper pulp industries, in detergent compositions, in the  
 CC textile industry, to improve the feed value of animal feedstuffs, and in  
 CC the food industry

XX SQ Sequence 1134 BP; 392 A; 228 C; 223 G; 291 T; 0 U; 0 Other;

Query Match 99.8%; Score 958.4; DB 6; Length 1134;

Best Local Similarity 99.9%; Pred. No. 4.5e-276;

Matches 959; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAGTTGCTCATCTGATCTATCTTAACTAATCTTTTAGTACAGGCAATA 60

Db 82 ATGAGCAAGAAAAGTTGCTCATCTGATCTATCTTAACTAATCTTTTAGTACAGGCAATA 141

QY 61 TATTTGTAGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAAAATACCTCATCTACA 120

Db 142 TATTTGTAGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAAAATACCTCATCTACA 201

QY 121 CCACCCCAACACACACTTTCCCTACCAAGTTCTCAAGATTAGATACCTCGATGACGGT 180

Db 202 CCACCCCAACACACACTTTCCCTACCAAGTTCTCAAGATTAGATACCTCGATGACGGT 261

QY 181 GAGTGCCAGGAGCTCCTATTGATAGGATGGTGGGAACCCAGAAATTTACATTGAA 240

Db 262 GAGTGCCAGGAGCTCCTATTGATAGGATGGTGGGAACCCAGAAATTTACATTGAA 321

QY 241 ATAACTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACATTTAAC 300

Db 322 ATAACTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACATTTAAC 381

QY 301 AGCGGGCTCCTTCACTACGTCCAAACACTTGACACATTTGCTTGAGGATAGAGTAAT 360

Db 382 AGCGGGCTCCTTCACTACGTCCAAACACTTGACACATTTGCTTGAGGATAGAGTAAT 441

QY 361 TGGTGTCATGATACCCCGAAATATTTCTATGAAACAAGCCATGGAATGCAAACTACGCA 420

Db 442 TGGTGTCATGATACCCCGAAATATTTCTATGAAACAAGCCATGGAATGCAAACTACGCA 501

QY 421 ACTGATGCCCAATACCATTTACCCAGTAAAGTTTCAAACTAACAGACTTCTATCTAAC 480

Db 502 ACTGATGCCCAATACCATTTACCCAGTAAAGTTTCAAACTAACAGACTTCTATCTAAC 561

QY 481 ATCTCCTATAAACTTTGAGCCCAAGAACCGGCTGCCAATTTAACTTCGCAATAGAACTCTGG 540

Db 562 ATCTCCTATAAACTTTGAGCCCAAGAACCGGCTGCCAATTTAACTTCGCAATAGAACTCTGG 621

QY 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAGAGAACTATGATA 600

Db 622 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAGAGAACTATGATA 681

QY 601 TGGATTTACTATGACGGATTACAACCGGCTGGCTCCAAAGTTTAAAGAGATTCTAGTCCCA 660

Db 682 TGGATTTACTATGACGGATTACAACCGGCTGGCTCCAAAGTTTAAAGAGATTCTAGTCCCA 741

QY 661 ATAAATGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720

Db 742 ATAAATGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGGAAGCAAACTGGT 801

QY 721 TGGGAGTATGTTGATTAAGAACCCCAATCAAAAGGGGAACAGTGACAAATTTCCA 780

Db 802 TGGGAGTATGTTGATTAAGAACCCCAATCAAAAGGGGAACAGTGACAAATTTCCA 861

QY 781 TACGAGCATTTTATAAGTGTTCGACCCCAACATTTCAAGCTTTACCAAAATTTACACAGACTT 840

Db 862 TACGAGCATTTTATAAGTGTTCGACCCCAACATTTCAAGCTTTACCAAAATTTACACAGACTT 921

QY 841 TACTTAGAGGAGCTGGAGATTGGAAGCTGAGTTTGGAAAGCCCAAGCACTACCTCCGCCAC 900

Db 922 TACTTAGAGGAGCTGGAGATTGGAAGCTGAGTTTGGAAAGCCCAAGCACTACCTCCGCCAC 981

QY	901	CTAGAGTGGTGGATCACAACATTAACACTTAACCTCTCTAGATAGACCTCTTTATTTCTTAA	960
Db	982	CTAGAGTGGTGGATCACAACATTAACACTTAACCTCTCTAGATAGACCTCTTTATTTCTTAA	1041
RESULT 6			
ID	ABA00497		
XX	ABA00497 standard; cDNA; 1134 BP.		
AC	ABA00497;		
XX	07-FEB-2003 (first entry)		
DT	P. furiosus EglA coding sequence.		
XX	Gene; variant; thermostable; cellulase; Cell12a; family 12; EglA;		
XX	glycosyl hydrolase; freeness; ink; coating; toner; colour; wood;		
KW	paper pulp; detergent; cellulose-containing textile; garment; lint;		
KW	fibrous crop; fruit; vegetable; grain; feed value; stability; solubility;		
KW	catalytic activity; cytotoxicity; ss.		
XX	Pyrococcus furiosus.		
OS			
XX	Location/Qualifiers		
EH	82..1041		
FT	/*tag= a		
FT	/product= "EglA"		
FT			
XX	US2002102699-A1.		
PN	01-AUG-2002.		
XX	23-OCT-2001; 2001US-00003759.		
PF	15-JUN-2000; 2000US-00594884.		
XX	15-JUN-2001; 2001WO-1S000012.		
PR	(PROK-) PROXARIA LTD.		
PA	Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;		
PI	WPI; 2002-749592/81.		
XX	P-PSDB; AAG79624.		
DR	Novel isolated nucleic acid encoding a polypeptide having thermostable		
XX	cellulase activity, useful for producing thermostable cellulose		
PT	polypeptide and as probes for isolating homologous sequences.		
PT	Disclosure; Page 14-15; 21pp; English.		
FS	This sequence encodes EglA, a variant family 12 glycosyl hydrolase		
XX	derived from P. furiosus. The amino terminal of EglA, amino acids 1-19,		
CC	constitute the hydrophobic domain. Residues 28-49 constitute the linker		
CC	moiety, with amino acids 50-319 forming the catalytic domain. EglA is		
CC	useful to improve freshness and to remove inks, coatings, toners and		
CC	colours from wood or paper pulp, in detergent compositions and to treat		
CC	cellulose-containing textiles and garments to improve the feel of the		
CC	fabric or to remove lint, in the treatment of fibrous crops, fruits and		
CC	vegetables or grains to improve feed values or to extract starches (e.g.		
CC	sugars) or other components of the crop, fruit, vegetable or grain being		
CC	treated. The variant glycosyl hydrolase polypeptide has improved		
CC	characteristics, such as increased stability (e.g. thermal stability,		
CC	detergent stability), increased solubility in aqueous solvents, increased		
CC	catalytic activity (e.g. specific activity, catalytic rate) and/or		
CC	reduced cytotoxicity relative to the native or full-length thermostable		
CC	cellulase, but retains the substrate specificity of the native or full-		
CC	length cellulase		
XX	Sequence 1134 BP; 392 A; 228 C; 223 G; 291 T; 0 U; 0 Other;		
XX	Query Match 99.8%; Score 958.4; DB 6; Length 1134;		
XX	Best Local Similarity 99.9%; Pred. No. 4.5e-276;		

DT 02-JUN-2003 (first entry)  
 DE pNOV4800 nucleotide sequence SEQ ID NO:56.  
 XX  
 KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;  
 KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;  
 KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;  
 KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ds.  
 XX  
 OS Pyrococcus furiosus.  
 OS Synthetic.  
 XX  
 PN WO2003018766-A2.  
 XX  
 PD 06-MAR-2003.  
 XX  
 PF 27-AUG-2002; 2002WO-US027129.  
 XX  
 PR 27-AUG-2001; 2001US-0315281P.  
 XX  
 FA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX  
 FI Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;  
 XX  
 DR WPI; 2003-268420/26.  
 XX  
 PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.  
 PT alpha-amylase, useful for producing plant to produce food products having  
 PT improved taste or fermentable substrates for ethanol.  
 XX  
 PS Disclosure; Page 114; 158pp; English.  
 XX  
 CC The present invention describes polynucleotides which encode processing  
 CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose  
 CC isomerase, or glucoamylase) that are optimized for expression in plants.  
 CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic  
 CC processing enzymes, which are activated under suitable conditions to act  
 CC upon the desired substrate. Also described are self-processing transgenic  
 CC plants and plant parts, e.g. grain, which express one or more of these  
 CC enzymes and have an altered composition that facilitates plant and grain  
 CC processing. Also described is a method (M) for converting starch to  
 CC starch-derived products in a transformed plant part (TPP), by activating  
 CC the starch processing enzyme contained in it. Transgenic grain is useful  
 CC for preparing maltodextrin. A transformed plant (TP) can be used to  
 CC produce food products having improved taste and to produce fermentable  
 CC substrates for ethanol and fermented beverages. (M) eliminates the need  
 CC to mill or physically disrupt the integrity of plant parts prior to  
 CC recovery of starch-derived products. The present sequence represents a  
 CC pNOV4800 nucleotide sequence (Amy32B signal sequence with EGLA), which is  
 CC given in the exemplification of the present invention  
 XX  
 SQ Sequence 978 BP; 206 A; 371 C; 240 G; 161 T; 0 U; 0 Other;  
 Query Match 55.9%; Score 536.8; DB 7; Length 978;  
 Best Local Similarity 74.4%; Pred. No. 6.1e-150;  
 Matches 676; Conservative 0; Mismatches 232; Indels 0; Gaps 0;  
 QY 52 CAGGCAATATATTTGTAGAAAAGTATCATACCTCTGAGGACAGTCAACTCAAAATACC 111  
 DB 70 CATCAATCTACTTCTGTTGAGAGTACACACCTCCGAGGACAAAGTCCACTCAACACC 129  
 QY 112 TCATCTACACACCCCAACACTTCCACTACCAAGTTCTCAAGATTAGATACCCCT 171  
 DB 130 TCCCTCCACCCCGCGAGACACACCTCTCCACCAACCAAGTGTCAAGATCCCGTACCCG 189  
 QY 172 GATGACCGTGTGCGGAGGAGTCCCTATTGATGAGGATGTGATGGACCCAGATTC 231  
 DB 190 GACACCGTGTGTTGGCGCGCGCCCGGATCGAACGAGACCGGCAACCCGGAGTTC 249  
 QY 232 TACATTGAAATTAACCTATGGAACATCTTTAATGCTACTGATTTGCTGAGATGAGGTAC 291  
 DB 250 TACATCGATCAACCTCTGGAACATCTTCAACGCCACCGGCTTCGCCGAGATGACCTAC 309  
 QY 292 AATTTAACAGCGCGCTCTTCACTACGTCCCAACACTTGACAACTTGCTTTGAGGAT 351  
 DB 310 AACCTCACTAGTGGCGTGTCTCCACTACGTGCGAGAGTCCGACACATCTGCTCCGCGAC 369  
 QY 352 AGAAGTAATTTGGTGTGATGGATACCCCGAAATATTCTATGGAAACAAGCCATGGAATGCA 411  
 DB 370 CGCTCCAACTGGGTGACGCGCTACCCGGAATCTTCTACGGCAACAAGCGGTGGACGCGC 429  
 QY 412 AACTAGCAACTGATGGCCCAATACCATTAACCCAGTAAAGTTTCAACACCTTAACAGACTTC 471  
 DB 430 AACTAGCCACCGACGCGCCCGATCCGCTCCGTCACAGTGTCCACACTCACCGACTTC 489  
 QY 472 TATCTAACAAATCTCTATTAACCTTGAGCCCAAGACGCGCTGCCAAATTAACCTTCGCAATA 531  
 DB 490 TACCTCACCATCTCTCTACAAAGCTCGAGCGGAAACGGTCTCCCGATCAACTTCGCGCATC 549  
 QY 532 GAATCTCTGTTAACGAGAGAACTGGAGAACACAGGAATTAACAGCGATCAGCAAGAA 591  
 DB 550 GAGTCTCTGGCTACCCCGAGGCTGGCGCACACCGGCATCACTCCACGAGGAGGAG 609  
 QY 592 GTAATGATATGGAATTTACTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATT 651  
 DB 610 GTGATGATCTGGATCTACTACGAGCGCTCCAGCGCGCGGGCTCCAAAGTGAAGGATC 669  
 QY 652 GTAGTCCCAATTAATAGTTAACCGAACACCAAGTAATGCTACATTTGAAGTATGGAAGCA 711  
 DB 670 GTGGTGCCGATCATGTGAACGGCACCCCGGTGAACGCCACCTTCGAGGTGTGGAAGGCC 729  
 QY 712 AACATTGGTGGAGTATGTTCATTTAGAAATAAAGACCCCAATCAAGAGGAGGAAACAGTG 771  
 DB 730 AACATCGCTGGAGTACGTGGCTTCGCAATCAAGACCCCGATCAAGAGGCGGACCGTG 789  
 QY 772 ACAATTCCATACGGAGCATTTAATAGTTGTGAGCAACATTTCAAGCTTACAAATTAC 831  
 DB 790 ACCATCCGTACGGCGCTTCTATCTCCGTGGCGGCAACATCTCTCTCCCTCCGCAACTAC 849  
 QY 832 ACAGAACTTTACTTAGAGACGCTGGAGATTGGAATGAGTTTGGAAACGCCAAGCACTACC 891  
 DB 850 ACGAGAGTACTCTGAGACGCTGGAGATCGGACCGGATTCGGACACCCCGTCCACCC 909  
 QY 892 TCGGCCCACTAGTGGTGGATCAACAACATPAACACTAATCTCTAGATAGACTCTTT 951  
 DB 910 TCGGCCCACTGAGTGGTGGATCAACAACATCAACCTCACCCCTCGACCCCGCTC 969  
 QY 952 ATTTCCTA 959  
 DB 970 ATCTCCTA 977  
 RESULT 8  
 ACC44577  
 ID ACC44577 standard; cDNA; 903 BP.  
 XX  
 AC ACC44577;  
 XX  
 DT 02-JUN-2003 (first entry)  
 XX  
 DE Pyrococcus furiosus EGLA nucleotide sequence SEQ ID NO:56.  
 XX  
 KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;  
 KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;  
 KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;  
 KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ss.  
 XX  
 OS Pyrococcus furiosus.  
 XX  
 FH Location/Qualifiers  
 CDS 1..903  
 FT /\*tag= a  
 FT /product= "EGLA"  
 FT /transl\_except= (pos:781..783,aa:1eu)  
 FT /note= "no start codon given"  
 XX



XX Claim 3; Fig 1A; 164pp; English.

XX This genomic DNA sequence from archaeal bacterial AEP11a codes for a

XX thermostable endoglucanase (see AAW34985) that is able to degrade

CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic bonds in

CC cellulose. AEP11a was discovered in a shallow marine hydrothermal vent.

CC It grows optimally at 85 degC and pH 6.5. The DNA (deposited in pOET as

CC ATCC 97516) was isolated from a genomic DNA library of AEP11a by

CC screening for clones having endoglucanase activity. It can be used in the

CC recombinant production of the endoglucanase and as a probe to identify

CC similar sequences. Other polynucleotides (see AAW34194-216) encoding

CC endoglucanases (see AAW34986-W3508) having homology to the AEP11a

CC endoglucanase are also claimed. The endoglucanases can also be used to

CC degrade cellulose for the conversion of plant biomass into fuels and

CC chemicals, for use in detergents, textiles, animal feed, waste treatment,

CC and in the fruit juice and brewing industries for the clarification and

CC extraction of juices. (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 2529 BP; 702 A; 719 C; 638 G; 470 T; 0 U; 0 Other;

SQ Query Match 12.1%; Score 115.8; DB 2; Length 2529;

Best Local Similarity 66.8%; Pred. No. 8.8e-24;

Matches 165; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 92 ACAAGTCACTTCAAACTCTACATCTACACACCCCAACACACATTTCCACTACCAAGG 151

DB 2276 AGACTTCAACCCACTACACACCACTTCAACCCGCGACACCCGCGCACTGCTCAGGACG 2335

QY 152 TTCTCAAGATTAGATACCTGATGACGGTGAGTGCGCCAGGAGTCTCTATTGATAAGGATG 211

DB 2336 TAATRAGCTCAGTACCGGACGATGGCGAGTGCGCCGAGGCCCAATTTGACAGGATG 2395

QY 212 GTGATGGGAACCCAGAAATTTCAATTTGAATAAACCCTATGGAACATTTCTTAATGCTACTG 271

DB 2396 GAGACGGAAACCCAGAGTTCTACATAGAAATAAACCCTGGAACATCTAGCGGCTGAAA 2455

QY 272 GATTGCTGAGATGACGTACGATTTAACCCAGCGCGCTCTTCACTAGTCCCAACAACTTG 331

DB 2456 GCTACGGCGAGATGACCTTACAACTTGACGACGGGGTTCTCCACTAGTCCAGGCCCTGG 2515

QY 332 ACAACAT 338

DB 2516 ATAGTAT 2522

RESULT 10

AAV40651

ID AAV40651 standard; DNA; 894 BP.

XX AC AAV40651;

XX 17-OCT-2003 (revised)

XX 26-OCT-1998 (first entry)

DE Endo-1,4-beta-glucanase gene in plasmid pMB447A.

XX Endoglucanase; cellulase; thermostable enzyme; textile; biopolishing;

KW stone-washing; saccharification; feedstuff; coffee; ss.

XX Bacillus sp.

OS Dictyoglomus; sp.

OS Chimeric.

XX Key Location/Qualifiers

FT sig\_peptide 1..87

FT /tag= a

FT /note= "Bacillus sp. signal peptide"

FT mat\_peptide 88..894

FT /tag= b

FT /note= "Dictyoglomus sp. endoglucanase gene"

XX W09828410-A1.

XX 02-JUL-1998.

XX 19-DEC-1997; 97WO-DK000583.

XX 20-DEC-1996; 96DK-00001483.

XX (NOVO ) NOVO-NORDISK AS.

PI Schuelein M, Bjornvad ME, Norrevang IA;

XX WPI: 1998-377642/32.

DR P-PSDB; AAW57777.

XX New endo-1,4-beta-glucanase active at high temperature and wide pH range

FT - useful, e.g. for improving properties of cellulosic textiles, e.g.

PT reduce pilling and improve softness.

XX Example 3; Page 32-33; 53pp; English.

XX This DNA sequence, from plasmid pMB447A, comprises a coding sequence for

CC a signal peptide (see AAV40648) of a Bacillus sp. and for mature

CC thermostable endoglucanase of Dictyoglomus sp. DSM 6262. The derived

CC protein sequence is provided in AAW57777. The open reading frame is

CC placed under the transcriptional control of the SPAC promoter in pMB447A.

CC The signal peptide portion of the construct directs the endoglucanase to

CC the exterior of host Bacillus sp. cells. Recombinant endoglucanase was

CC expressed in Bacillus subtilis. The enzyme shows optimum activity at a

CC temperature above 85 degC, and exhibits an activity towards CM-cellulose

CC at 70 degC and pH 10 that is higher than 50% relative to the activity at

CC 70 degC and the optimum pH. A Bacillus subtilis strain harboring an

CC expression plasmid encoding the thermostable endoglucanase cloned from

CC Dictyoglomus sp. DSM 6262 is deposited as DSM 11903. The endoglucanase

CC can be used to improve properties of cellulosic textiles, e.g. to reduce

CC pilling, improve softness, to provide a stone-washed look to denim, for

CC industrial cleaning, e.g. of ultrafiltration membranes and pipes, in heat

CC -extrudable polymers (to increase degradability), in conversion of

CC biomass to sugars, in production of ethanol (particularly preliquefaction

CC of grain), for improving digestibility of fodder grains, and in

CC production of instant coffee or other similar extraction processes

CC (increasing capacity of the extraction column) (all claimed). Also

CC provided are DNA constructs encoding the endoglucanase, recombinant

CC expression vectors and host cells (preferably Bacillus, Dictyoglomus or

CC Saccharomyces species or a filamentous fungus). (Updated on 17-OCT-2003

CC to standardise OS field)

XX SQ Sequence 894 BP; 324 A; 167 C; 162 G; 241 T; 0 U; 0 Other;

Query Match 11.9%; Score 114.2; DB 2; Length 894;

Best Local Similarity 51.3%; Pred. No. 1.7e-23;

Matches 291; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

QY 230 TCTACATTGAAATAAACCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACCT 289

DB 164 TCACCTTAGAATCAACTTTTGGATATTGCAAACTATGAGGAATAATCATGATGGCAT 223

QY 290 ACAATTTAACCGCGCGCTCTCTACGTCCCAACAACTTCACAACTTGTCTTGAAGG 349

DB 224 TTATATAAGAGAAGATACCTGTTGAATATTATGCCGACATAAAACATAGTACTTAAGG 283

QY 350 ATAGAAGTAATTGGGTGCATGATACCCGCAATATTCTATGGAACAGCCATGGAATG 409

DB 284 ATAAAAAATTCATGGGTACATGATATCTCTGAAGTCTACTATGGGTACAAACCATGGGTG 343

QY 410 CAAACTACGCAACTGATGGCCCAATACCAATTACCCAGTAAAGTTTCAAACTTAACAGACT 469

DB 344 GCCATGGGAATTCATTGAGAAATTAGCTCTTCTTAAAGGTATCAGAAATTTCCAGAGC 403

QY 470 TCTATCTAACAACTCTCTTATAAACTTGAGCCCAAGAACCGGCTGCCAATTAATCTCGCAA 529

DB 404 TTCTCTTCAATCTAAAAATACACATATGTTAGTACGAGAAATCTTCTATAAATTTGCTA 463

QY 530 TAGAATCCTGGTTAACGAGAGAAGCTTGGAGAACCAACAGGAATTAACAGCGATGAGCAAG 589



```

Db 464 TGGAAACATGATACAAAGAACCTATCAGAAAACCG---TTACTTTCAGGGGATATAG 520
QY 590 AAGTAATGATATGGATTACTATGACGATTACACCGGCTGCTCCAAAGTTAAAGAGA 649
Db 521 AGATGATGGTATGGCTATATGCTATAGACTTCTCCGAGGGGAAAGTAGGAGAAG 580
QY 650 TTGTAGTCCCAATATAGTTAAACGGAACACAGTAAATGCTTACATTTTGAAGTATGGAAG 709
Db 581 TAAAAATACCTATCATCTCTAAACGGTAATCAAAAAGACATTATCTGGAAGTATATCTTT 640
QY 710 CAAACATTTGTTGGGAGTATGCTTGAATTAAGATAAGACCCCAATCAAGAGGACAG 769
Db 641 CCCCTATGAGCTGGGACTACGTGGCTATATAATCAAAAAGAAAATATCTTCAAGGACAG 700
QY 770 TGACAAATTCATACGAGCAATTTATAA 796
Db 701 TAAAAATACCAATAAATGAATTTTGA 727

RESULT 11
AAV47544
ID AAV47544 standard; DNA; 894 BP.
XX
AC AAV47544;
XX
DT 18-NOV-1998 (first entry)
XX
DE PMB447A plasmid DNA sequence encoding an endo-beta-1, 4-glucanase.
XX
KW Pyrococcus furiosus endo-beta-1, 4-glucanase; cellulolytic activity;
KW textile industry; cellulosic fibre; in industrial cleaning process;
KW sugar; instant coffee; oil industry; hydrocolloid cellulose derivative;
KW drilling; chimeric; ss.
XX
OS Synthetic.
XX
SS Dictyoglomus sp.
XX
FH Key
CDS Location/Qualifiers
FT 1..894
FT /*tag= a
FT /product= "Dictyoglomus species endo-beta-1, 4-glucanase"
FT sig_peptide 1..87
FT /*tag= b
FT /note= "Signal peptide encoding sequence derived from
FT Bacillus species"
FT mat_peptide 88..891
FT /*tag= c
XX
PN WO9833895-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-DK000039.
XX
PR 31-JAN-1997; 97DK-00000114.
XX
PR 11-JUL-1997; 97DK-00000853.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Andersen L, Bjornvad ME, Schuelein M;
XX
WPI; 1998-437450/37.
DR P-FSDB; AAW29730.
XX
PT Isolated endo-beta-1,4-glucanase - used for e.g. treating cellulosic
PT fibres or polymers, feed production or in oil industry for enhancing oil
PT recovery.
XX
PS Disclosure; Page 31; 56pp; English.
XX
SS The present sequence represents the region of the pMB447A plasmid
CC encoding a Dictyoglomus species endo-beta-1, 4-glucanase (5G). This

```

```

CC vector was given as an example of an expression vector that could be used
CC in the method of the invention. The invention provides for a Pyrococcus
CC furiosus endo-beta-1, 4-glucanase encoding DNA (AAV47540) isolated from a
CC P. furiosus genomic DNA library. The invention provides for an enzyme
CC composition having EG activity which has optimum activity at a
CC temperature of at least 90 deg. C. The EG enzyme exhibits cellulolytic
CC activity at extremely high temperatures in a very broad pH range.
CC Therefore it is claimed to be useful for, e.g. in the textile industry
CC for improving the properties of cellulosic fibres or fabric, for
CC providing a stone-washed look of denim; in industrial cleaning processes;
CC in the conversion of biomass to sugars; in the production of instant
CC coffee or similar extraction processes or in the oil industry for
CC degradation of aqueous solutions of hydrocolloid cellulose derivatives
CC used in drilling
XX
SQ Sequence 894 BP; 324 A; 167 C; 162 G; 241 T; 0 U; 0 Other;
Query Match 11.9%; Score 114.2; DB 2; Length 894;
Best Local Similarity 51.3%; Pred No. 1.7e-23;
Matches 291; Conservative 0; Mismatches 273; Indels 3; Gaps 1;
QY 230 TCTACATTGAAATAAACCTATGGAACATTTCTTAATGCTACTGGATTCTCTGAGATGAGT 289
Db 164 TCACCTTAGACTCAACTTTTGGAAATTTGCAAACTATGAAAGAAATACATGGATGGCAT 223
QY 290 ACAATTTAACACGCGGCTCCTTCACTAGCTCCACCACTTCACACCACTTGCTCTGAGG 349
Db 224 TTTATAAGAGAAGATACTGTTGAATATTATGCGGCATATAAACAATAGTACTTAAGG 283
QY 350 ATAGAAGTAATTTGGGTGATGATACCCCGAAATATTCTATGGAACAAGCCATGGAATG 409
Db 284 ATAAAAATTCATGGGTACATGATATCTCTGAAGTCTACTATGGGTACAAACCATGGGCTG 343
QY 410 CAAACTACGCACTCATGCCCCAATACCATTACCCAGTAAAGTTTCRAACCTACAGACT 469
Db 344 GCCATGGGAATTTCCATTGAGAAATTAGCTCTTCTTAAAGAGGTATCAGAAATTCAGAG 403
QY 470 TCTATCTACAATCTCCTATAAACTTGAGCCCAAGACGCGCTGCCAATTAACITTCGCA 529
Db 404 TTCTCTTCAATCTAAATACACATATGTACGAGAAGAACTTCCTATAAATTTTGCTA 463
QY 530 TAGAATCTCTGGTTACGAGAGAGCTTGGAGAACAAACAGGAATTAACAGCGATGAGCAAG 589
Db 464 TGGAAACATGGAATAACAAAAGAACCCCTATCAGAAAACCG---TTACTTTCAGGGGATATAG 520
QY 590 AAGTAATGATATGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAAGAGA 649
Db 521 AGATGATGGTATGGCTATATGCTAATAGACTTTCTCTGAGGGGAAAGTAGGAGAAG 580
QY 650 TTGTAGTCCCAATAATAGTTAAACGGAACACCAAGTAAATGCTACATTTTGAAGTATGGAAG 709
Db 581 TAAAAATACCTATCATCTCTAAACGGTAATCAAAAAGACATTATCTGGAAGTATATCTTT 640
QY 710 CAAACATTTGTTGGGAGTATGTTGCAATTTAGATAAGACCCCAATCAAGAGGAAACAG 769
Db 641 CCCCTATGAGCTGGGACTACGTGGCTATATAATCAAAAAGAAAATATCTTCAAGGACAG 700
QY 770 TGACAAATTCATACGAGCAATTTATAA 796
Db 701 TAAAAATACCAATAAATGAATTTTGA 727

RESULT 12
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;

```



KW gene; ds.  
XX Oryza sativa.  
OS WO2003000898-A1.  
PN 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-173290/17.  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 27; SEQ ID NO 5263; 899pp; English.  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
SQ Query March 4.9%; Score 47.4; DB 7; Length 2000;  
Best Local Similarity 10.4%; Pred. No. 0.0023;  
Matches 75; Conservative 318; Mismatches 324; Indels 4; Gaps 2;  
QY 10 AAAAGTTCGTCATCTATCTTAAACATCTTTAGTACAGGCAATATTTGTA 69  
DB 944 RWRWAWCYCCWKKWKTSCWKKYRTWCSWTWKGARVAYAYRRRTYKWSR 885  
QY 70 GAAAGTATCATCTCTGAGGACAAGTCACTCAATCACTCTACACACCCAA 129  
DB 884 RMYTMTKAWTWTCTCMKAKWYMATGWATWWRWYTYCYANTCAKCKYKAMTKW 825  
QY 130 ACAACACTTTCCACTACCAAGTTCTCAAGATTAGATACCTGATGACGGTGGCCA 189  
DB 824 TTWACAWATSWRWRWAGWKKYKRYWWRWRCWAGWARKWRSYRWKKY - 767  
QY 190 GGAGCTCTATGTATAAGGATGGTATGGGACCCAGAAATCTACATGAAATAACCTA 249  
DB 766 ATRYWKKWAMTWWSWRWKSXWWSGGMWRWSAWRYCSRMKCAKTKYASSARWTKRA 707  
QY 250 TGGACACTTCTATGCTGCTGATGCTGAGATGACGTACATTTAAACGCGCGTC 309  
DB 706 KRSYRRRRWYKKGWYRYRYRWSRMTARSKRRKRWAGASMKSWWYRGARSMWY 647  
QY 310 CTTCACTAC--GTCCAAACATTCGACACATTTGCTTGGGATAGAAATTTGGTGC 367  
DB 646 SKYSCSAKCKKTRTYMTSSYSTMGYSSYKSWSTSKWSYMGKVTCTMTYISMKGST 587  
QY 368 ATGATATCCCGCAATATCTATGGAACACGACCATGGAATGCAAACTACGCACTGAT 427  
DB 586 RSKGRWSGSMRYRWKWKRYWYKWKCTWRRCMYRNGYTYMTTSRSMYTG 527  
QY 428 GCCCAATACCATTCACCAAGTAAAGTTTCAACACCTTAACAGACTTCTATCTAACATCTCCT 487

DB 526 RYKARYTSKRYYMYKRYKYYWYMYKCSYMYRYGYKACKKCYAMCWKAAYSGM 467  
QY 488 ATAAACTTTCAGCCCAAGAACGCGCTCCCAATTAATTCGCAATAGATCTGGTTAACGA 547  
DB 466 MMVYRYKYSKWRMSTKYWWSMWYKCRSMKYGAKGYCKWWTYCSYGMKWYVNGSY 407  
QY 548 GAGAAGCTTGGAGAACACAGCAATTAACAGCGATGAGCAAGATAATGATGATGATT 607  
DB 406 KYRCYKYMRYMYKGMWYMYYSAYSSMMTYYIYAKYKRYKRRGTMSYKSYKXK 347  
QY 608 ACTATGACGGATTACAAACGCGCTCGCTCCAAAGTTAAGGAGATTGTAGTCCCAATAATAG 667  
DB 346 YCTWYCMKRCYRWKWKTKYKCYCWRYATCYWCCYKRGWYSRSMRTAG 287  
QY 668 TTAACGGAAACACAGTAATGCTATTTGAAGTATGGAAGCAACATTTGGTGGAGT 727  
DB 286 KWRMSWSRWSRYSYMYKWKYKWKYSYMSYMSYMSYMSYMSYMSYMSYMSYMSYMSY 227  
QY 728 A 728  
DB 226 M 226  
RESULT 13  
AAAX80187  
ID AAX80187 standard; DNA; 999 BP.  
XX AAX80187;  
AC AAX80187;  
DT 17-AUG-1999 (first entry)  
XX Sulfolobus solfataricus endo-beta-1,4-glucanase Celsa encoding DNA.  
DE Sulfolobus solfataricus endo-beta-1,4-glucanase Celsa; CelB;  
XX Sulfolobus solfataricus; endoglucanase; Celsa; CelB;  
KW endo-beta-1,4-glucanase; ss.  
XX Sulfolobus solfataricus.  
FH Key Location/Qualifiers  
FT CDS 1..999  
FT sig\_peptide /\*tag= a  
FT mat\_peptide /\*tag= b  
FT /\*tag= c  
XX DK9900097-A.  
XX 12-JAN-1999.  
XX 12-JAN-1999; 99DK-00000097.  
XX 12-JAN-1999; 99DK-00000097.  
XX (NOVO ) NOVO-NORDISK AS.  
XX WPI; 1999-279376/24.  
DB P-PSDB; AAY17865.  
XX Novel endonuclease - produced by Sulfolobus solfataricus.  
XX Claim 9; Page 26-27; 36pp; English.  
XX The present sequence encodes a Sulfolobus solfataricus endoglucanase,  
CC specifically endo-beta-1,4-glucanase, designated Celsa. The new endo-beta-  
CC 1,4-glucanase is produced by Sulfolobus solfataricus ATCC 35092  
XX SQ Sequence 999 BP; 317 A; 212 C; 183 G; 287 T; 0 U; 0 Other;  
Query Match 4.9%; Score 47.2; DB 2; Length 999;  
Best Local Similarity 52.0%; Pred. No. 0.002;  
Matches 106; Conservative 0; Mismatches 98; Indels 0; Gaps 0;



Db 27270 AGCAAAATTGGCGCTGATACCAAGAGCCTTTAACCAAGAGCTTTAATGAGCAAGTCAA 27211  
Qy 324 ACAACTTGACAAACATGTGCTTGAGGGATAGAGTAATT 361  
Db 27210 AGGGCTTGGCAGCAGCGGCAAGGGAGATATAAAGCAATT 27173

Search completed: July 1, 2004, 21:58:15  
Job time : 609 secs





[illegible]

RESULT 4  
G90291  
endoglucanase precursor [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: G90291  
R:She, Q.; Singi, R. K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.;  
Jong, I.; Jeffries, A. C.; Rozera, C. J.; Medina, N.; Peng, X.; Thi-Ngoc, H. P.;  
arrett, R. A.; Ragan, M. A.; Sengen, C. W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A:description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: G90291  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-332 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3814564; PIDN:AAK41590.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SS01354

Query Match	25.3%;	Score	430;	DB 2;	Length	332;
Best Local Similarity	32.2%;	Fred. No.	4.5e-25;			
Matches 111;	Conservative	60;	Mismatches	124;	Indels	50;
Gaps	15					
Qy	4	KKFFVIVSLITLLV	-----QAIFYFEKYHTSEDKSTNTSSTP	-----	PQTL	46
Db	3	KLVIVLPVIVIAIGVMGGIIVLHQQLSVKPVTTTFSTSTTTNAITTTVTQTVT	-----			62
Qy	47	STTKVLKIRYPDDGSGWPGAPIDKQDGNPEYIETINLWLNILNATGPAETYNLTSGVLHV	-----			106
Db	63	SITSYNQLIYVTSASSTPVPYLNASTISPFYLSVNNWNAKNTNGNTVWFNPDLATLSV	-----			122
Qy	107	VQQLDNVILDRSNWVHGYPFIYCKNPKWANYATDGPILPSPKVSNLT	-----DFYLTIS	162		
Db	123	SPNLTVQ---KPLEWTCNGYPELIYGRKPDWTAYA-GNI-FPMRIGNMTPEWVSFYINLT	176			
Qy	163	YKLEPNKGLPINF--AISWITREAWR---TTGINSDEQEVMIWIYDGLQPAAGSKVEI	217			
Db	177	-KLDPS-----INFDSADVAIVRPOIAFSPGTAPGNGDIEIWMVLFSONIQAPEGQVKV	231			
Qy	218	VPEIIWNGTPVNATPEVWKANI-----GWEYVAFRIKTP-----IKEGVTPIPYGAFISVAA	269			
Db	232	VPEIIVTNTTNNATFOVWEMKSVPGWVEYIAFR---PDGKVTNGVYVSYPENFLFKALS	288			

QY      270 NISSLENNVTELYLEDVEIGTFFGT-PSTISAHLEWNIINILTEL 313  
         :  
Db      289 NFISY-NITNYILTDFWEGCTGWTNGTAFTSVTSFSETLL 332

RESULT 5  
H90425  
hypothetical protein celB [imported] - Sulfolobus solfataricus  
C/Species: Sulfolobus solfataricus  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C/Accession: H90425  
R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.;  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A/Description: Sulfolobus solfataricus complete genome.  
A/Reference number: A99139

A;Accession: H90425  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-322 <KUR>  
A;Cross-references: GB:AE006641; NID:gl3815835; PIDN:AAK42663.1; GSPDB:GN00155  
C;Genetics:  
A;Gene: celB

Query Match 13.0%; Score 221.5; DB 2; Length 322;  
Best Local Similarity 27.3%; Pred. No. 8.1e-10;  
Matches 89; Conservative 52; Mismatches 132; Indels 53; Gaps 17;

QY 5 KFIIVSLTILLVQAIIVE--KTH--TSEDKSTNSSTP-----POTTSTTKVLKIRY 56  
DB 3 KLIPFVVIIIVLGIIVSIEFGKPHQNASLTRSTERTFLFAHNRPFSL----- 52

QY 57 PDDGEWPGAPDK-----DGDGNPEFYIEINLWNLINATGFAEMTYNLTSGVLHVQOLDN 112  
DB 53 ---GNYSNSADAILNSSTNATLMVSPFLWNLGYALGNVNTINI--NYLHVAINLSQ 107

QY 113 IVLDRSNVHGYEPIFYGNKP--WNANYATDGP--IPLPSKVSNLTDFTYLTISYKLEPKN 169  
DB 108 I-SKISSNVVDGYPLMGYQBLWPFMYRTTQLOFLSLPMIVLRLPNFYILNYSVYLIN 166

QY 170 GLPNFAIESLWTEARWTTGINSDEQVMIWYD--GLQPAGSKVKIIVVPLVNGT 226  
DB 167 GSIDDFSYDIWLSNP--NITSLOQDPEIMLWMNENLSHTFYIYVGNMSIPTLNGK 225

QY 227 PVNATFEVW-----KANIGWEYVAF--RIKTPKEGTVTIPY-----GAFISVA-ANIS 272  
DB 226 IENLSWEVYVLPRTGSAN--GWTGVYFLSPLEKPAEFGVPIGYILKNMGSYIEKAGVNIY 284

QY 273 SLPNYTELYLEDVIGTEGTPPTTS 298  
DB 285 NVNTY---YLDIAIQVGEFSDNQTA 307

RESULT 6  
JU0328  
cellulase (EC 3.2.1.4) precursor - Erwinia carotovora subsp. carotovora  
N;Alternate names: endo-1,4-beta-glucanase; endoglucanase  
C;Species: Erwinia carotovora subsp. carotovora  
C;Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 08-Oct-1999  
C;Accession: JU0328  
R;Saarilahti, H.T.; Henriksa, B.; Palva, E.T.  
Gene 90, 9-14, 1990  
A;Title: CelS: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora  
A;Reference number: JU0328; MUID:90337352; PMID:2379837  
A;Accession: JU0328  
A;Molecule type: DNA  
A;Residues: 1-264 <SAA>  
A;Cross-references: GB:M32399; NID:gl48389; PIDN:AAA24817.1; PID:gl48390  
A;Experimental source: strain SCC3193  
A;Note: part of this sequence, including the amino end of the mature protein, was confirmed by N-terminal sequencing  
C;Genetics:  
A;Gene: celS  
C;Function:  
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose  
A;Pathway: cellulose degradation  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation  
F;1-32/Domain: signal sequence #status predicted <Sig>  
F;33-264/Product: cellulase #status experimental <NAT>

Query Match 6.6%; Score 112.5; DB 2; Length 264;  
Best Local Similarity 21.8%; Pred. No. 0.2;  
Matches 55; Conservative 38; Mismatches 90; Indels 69; Gaps 14;

QY 68 DKDGDGHPFEFYIEINLWNLINATGFAEMTYNLTSGVLHVQOLDNVLDRSNW----- 121  
DB 40 DKLYFGNNKYVLFNPNVAGKDIKGWQQTIFY-----NSPISGMWVHWPSSST 86

QY 122 --VHGYPEIFYGNKPMNANYATDGPILPLPSKVSNLTDFTYLTISYKLEPKNGLPINFAIES 179

DB 87 HSKAYPSLVSG-WHWTAGYTENSLPI--QSSNKSITSNTVYSIKATG--TYNAAVDI 141  
QY 180 WL---TREAWRTTGINSDEQVMIWYDGLQPAGSKVKIIVVPIVNGTPVNATFEVWK 236  
DB 142 WHHTDKANWD-----SSPTDELMIMLNDTNAGPAGDYIETVFLG-----DSSWVFK 189

QY 237 ANI-----GWEYVAFRIKTPKEGTVTIPYGAFTSVANISSLPNY-----TELY 281  
DB 190 GWINADNGGGMVFSF-VHT---SGT-----NSASLNIHRFTDYLVTQKQMSDEKY 237

QY 282 LEDVEIGTF-FG 292  
DB 238 ISSVEFGTIFG 249

RESULT 7  
JC2571  
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)  
N;Alternate names: endo-1,4-beta-glucanase; endoglucanase  
C;Species: Streptomyces rochei  
C;Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 22-Oct-1999  
C;Accession: JC2571; S34392  
R;Perito, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastronei, G.  
Gene 148, 119-124, 1994  
A;Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase  
A;Reference number: JC2571; MUID:95011642; PMID:7523249  
A;Accession: JC2571  
A;Molecule type: DNA  
A;Residues: 1-382 <PER>  
A;Cross-references: EMBL:X73953; NID:g933391; PIDN:CAAS2139.1; PID:g933392  
A;Note: this cellulolytic strain was isolated from the gut of termites  
C;Genetics:  
A;Gene: eglS  
C;Function:  
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose  
A;Pathway: cellulose degradation  
C;Superfamily: bacterial cellulose-binding domain homology  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation  
F;1-37/Domain: signal sequence #status predicted <Sig>  
F;38-382/Product: endoglucanase #status predicted <NAT>  
F;279-380/Domain: bacterial cellulose-binding domain homology <BCB>  
F;280-379/Disulfide bonds: #status predicted

Query Match 6.1%; Score 103.5; DB 2; Length 382;  
Best Local Similarity 25.1%; Pred. No. 1.6;  
Matches 43; Conservative 29; Mismatches 70; Indels 29; Gaps 10;

QY 125 YPEIFYGNKPMNANYATDGP-IPLPSKVSNLTDFTYLTISYKLEPKNGLPINFAIESWLTR 183  
DB 94 YPSVFENG-----CHYTNCSPGTALPARISGISAPSSISYGF--VDNAVYNASYDIWLDP 146

QY 184 EAWRTTGINSDEQVMIWYDGLQPAGSKVKIIVVPIVNGTP--VNATFEVWKANIG 240  
DB 147 TP-RDGVN--RTEIMVFNVRVGOIQPGSQV-----GTASVAGRTWEVWGGNG 193

QY 241 W-EYVAFRIKTPKEGTVTIPYGAFTSVANISSLPNYTELYLEDVEIGTE 290  
DB 194 TNDVLSFVAPSANSSWSFDVFRVATVARGLAG----NDWYLTSTIQAGFE 240

RESULT 8  
T29630  
hypothetical protein K09E3.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29630  
R;Johnson, D.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid K09E3.  
A;Reference number: Z20655  
A;Accession: T29630  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA

A;Residues: 1-859 <JOH>  
A;Cross-references: EMBL:U41033; PIDN:AAA82378.1; CESP:K09E3.7  
C;Genetics:  
A;Gene: CESP:K09E3.7  
A;Introns: 19/3; 247/3; 428/3; 546/3; 664/3; 782/3

Query Match 6.1%; Score 103; DB 2; Length 859;  
Best Local Similarity 21.1%; Pred. No. 5.3;  
Matches 67; Conservative 41; Mismatches 115; Indels 94; Gaps 14;

QY 28 TSEDKSTNT-----SSTPPQTLST-----TKVLKIRYPDDGE 61  
DB 336 TLQDGTNDTNSNKRKAGSIBELLPDSTITQTSTPITAFISTTTPASTVPTMS 455  
QY 62 WFGAPIDXGDNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRS-- 119  
DB 456 FPSIP-----HIATLGNVAPINSLTSSLR--IDHQREIPASTLKRRSVS 498  
QY 120 -----NWVHGYPEIFYGN-----KP-----WNANYATDGPPLPSKVSNLTFD 157  
DB 499 DRMGLPCKRNLNPSVMSHVNGSTWEAKPLFPQWNLHPAMAPFPPTQSTPTSPF 558  
QY 158 YLTI-----SY-----KLEPPKGLPINFATIESWLTREAWRTTGINSDEQVMIWI 202  
DB 559 IPTPTSTLVRIMSFPNIHGTISFGNNAPINSRVSSLRAMEQQORTVTSASTLKQEA-VFE 617  
QY 203 YDGLQ-PAGSKVKEIVVPIVINGTPVNA-----TTEVVKANIGWYVAFRIKTIKEGTVT 258  
DB 618 MVNGLSYQRLNGTNGPIPTQFRGNSMEAKOSRPFVWN-----QHPAMPJSLPTQSTPPT 672  
QY 259 IPYGAFISVAANISLP 275  
DB 673 IP-AFIPTISTSM 687

## RESULT 9

T20015

hypothetical protein T13H10.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T20015; T24895

R;Coles, L.

submitted to the EMBL Data Library, January 1996

A;Reference number: Z19210

A;Accession: T20015

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-533 <WIL>

A;Cross-references: EMBL:Z68882; PIDN:CAA93112.1; GSPDB:GN00022; CESP:T13H10.1

A;Experimental source: clone C47E12

R;Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A;Reference number: Z19949

A;Accession: T24895

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-533 <W12>

A;Cross-references: EMBL:Z69361; PIDN:CAA93289.1; GSPDB:GN00022; CESP:T13H10.1

A;Experimental source: clone T13H10

C;Genetics:

A;Gene: CESP:T13H10.1

A;Map position: 4

A;Introns: 8/2; 62/3; 101/3; 203/3; 282/2; 355/3; 449/2; 490/3

## Query Match

Best Local Similarity 28.7%; Score 102.5; DB 2; Length 533;

Matches 51; Conservative 28; Mismatches 50; Indels 49; Gaps 15;

QY 96 TYNLTSGV-LHYVQQLDNIVLRDRSNWVHGYPEFYGNKPNWANYATDGPPLPSKVSNL 154

DB 227 TMVQAGLGLHYLHSL-QIHRDIAS-----RNCLYGN-----GQV-----KI 263

QY 155 TDFYLT---ISYKLEPKNGLPINFATIESWLTREAWRTTGINSDEQE-----VMIW-IYYD 205

DB 264 SDFGLSREGYSYRMNPHKKVPJR-----WLAPEVPR-TGFTPKDVFAYGVMCEVHD 317  
QY 206 GLQP-AGSKVKEIVVPIVINGTPVNNATFEVWKANIG---WEYVAFRIKTIKEGTVTI 259  
DB 318 GIEPYFGMKVAB-VLPRVQNG--YRMPFE--ANVPPAIVRFITVTRICAGAEERVTM 369

## RESULT 10

T21201

hypothetical protein F21D5.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T21201

R;Berks, M.

submitted to the EMBL Data Library, September 1995

A;Reference number: Z19389

A;Accession: T21201

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-713 <WIL>

A;Cross-references: EMBL:Z54271; PIDN:CAA91039.1; GSPDB:GN00022; CESP:F21D5.3

A;Experimental source: clone F21D5

C;Genetics:

A;Gene: CESP:F21D5.3

A;Map position: 4

A;Introns: 5/3; 41/3; 93/2; 180/1; 211/3; 252/1; 293/1; 332/3; 399/1; 459/3; 495/3; 628/3

Query Match 5.9%; Score 101; DB 2; Length 713;

Best Local Similarity 24.5%; Pred. No. 5.9; Indels 66; Gaps 15;

Matches 70; Conservative 28; Mismatches 122; Indels 66; Gaps 15;

QY 14 ILLVQAIYFEVKYHTS---EDKSTNTSSTPQTTLSTTKVLKIRYPDDGMPGAP---- 66

DB 376 IPVYFGYIE-YETDGLPEDXTTFLMH-PECTQNKKTVLNCPFQ---EMWAPNFTC 430

QY 67 --IDKDGDNPEFYIEINLNLNAT---GFAETYNLTSGVLHYVQQLDNIVLRDRSN 120

DB 431 ISYDKLENDPE-KTEA---ELQATQFNGGFEHFIN-----MHDSQMDQ----- 473

QY 121 WHGYPEIFYGNKPNWANYATDGPPLPSKVSNL-----DFYLTISYKLEPKNG 170

DB 474 -----FLFPQPMGIPYHGDGMTETSTSCQATCANNSDLDTTCRCFYHLEHKLN 525

QY 171 LPINFATIESWLTREAWRT-----TCINSDEQVMI--WYYDGLQAPAGSKVKEIVVPIV 223

DB 526 NIVQITLYNMLGGAMGTGYAHPFLHGHFVYMKVGPWSYNGSGFIDQMNQIDCP--- 582

QY 224 NGTPYNATEVVKAN--IGWEYVAFRIKTIKEGTVTIYGAFIGV 267

DB 583 -GRDVSCKGKRWKNDWLGGALEAWNTKNTKRTITLTPVGGYITI 627

## RESULT 11

S12610

cellulase (EC 3.2.1.4) precursor - Aspergillus aculeatus

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Aspergillus aculeatus

C;Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 21-Jul-2000

C;Accession: S12610; S14118; S40186; JQ0458

R;Ooi, T.; Shimmyo, A.; Okada, H.; Murao, S.; Kawaguchi, T.; Arai, M.

Nucleic Acids Res. 18, 5894, 1990

A;Title: Complete nucleotide sequence of a gene coding for Aspergillus aculeatus cellulase

A;Reference number: S12610; MUID:91016934; PMID:2216782

A;Accession: S12610

A;Molecule type: DNA

A;Residues: 1-237 <OOII>

A;Cross-references: EMBL:D00546; NID:g217818; PIDN:BA00435.1; PID:g217819

R;Ooi, T.; Shimmyo, A.; Okada, H.; Hara, S.; Ikenaka, T.; Murao, S.; Arai, M.

Curr. Genet. 18, 217-222, 1990

A;Title: Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from Aspergillus

A;Reference number: S14118; MUID:91064758; PMID:2249253

A;Accession: S14118





```
Query Match      5.8%; Score 99; DB 1; Length 1680;
Best Local Similarity 22.8%; Pred. No. 27;
Matches 66; Conservative 32; Mismatches 95; Indels 96; Gaps 13;

QY 81 INLWNI-----LNATGFAETYNLTGVLHYVQOLDNIVLRDRSNWVHGYPFIFG---131
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MGLWGLILLFLDKTWGQEQYVISAPKILRVGSSENVVIQ-----VHGYTEAFDALS 55
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 132 -----NKPNNYATDGPILPSPKVSNTLDFYL-TISYKLEPKN 169
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 LKSPYDKKVTSSGVNUSPENKQNAALLTLQNPVREESPVSHVLEVVSKHFSK 115
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 GLPINFAESWLTREAWRTIGI-----NSDEQEVMIWIIY--DGLQPA-----210
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 KIPITV-----NNGILFHTDKPVYTPDQSVKIRVYSLGDDLPKAKRETVLTF 163
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 -----GSKYKEIWPVPIVNG-----TPVNATEFVW--KANIG-----WEY 243
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 IDPEGSEV-DIVEENDYTGISFPDFKIPSPKPYGVMTIKANYKKDFTTGTAYFEIKEY 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 VAFRIKTKEGTVPYPGAFISVAANISSLPNYTELYLEDVEIGTFBG 292
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 VLPRESVIELERTFYGKFNKFEITVKARYFYFNKV-VPDAYVAFG 270
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
B99782
hypothetical protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B99782
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A99758; PMID:21311952; PMID:11418146
A:Accession: B99782
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-572 <KUR>
A:Cross-references: GB:BA000018; PID:gl3700117; PIDN:BA041416.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0195

Query Match      5.8%; Score 98.5; DB 2; Length 572;
Best Local Similarity 21.6%; Pred. No. 6.9;
Matches 69; Conservative 35; Mismatches 83; Indels 133; Gaps 16;

QY 6 FVIVS--ILTILLVOAIYFEVK---YHTSED---KSTNTSSTPPQTTLSTTKVKIRYP 57
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 FVNISFVITIALFTALYFNKIGWYRKEDIYSSNINFSLSPIKAITIQK---332
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 DDGEWPGAPIDKDGDN---PEFYIEINLWNLNA-TGFAETYNLTSGVLHYVQOLDNI 113
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 ---HW-----ISNKEVNLHKPPFYISYSISIFFLGGLIYFSQAYNQLANAVIFILIL-NT 384
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 VLDRSNWVHGYPFIFGYNKPNANYATDGPILPSPKVSNTLDFVLTISYKLEPKNGLPI 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 ITRD-----SPSAGTDFF-----397
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 NPAIESWLTREAWRTGINSDEQEVMIWIIY---YDGLQPAAGSKV-----KEIVVPI 221
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 -----TKSLRFSDRSKIGLYRMSNTDFKQIYDSKLSLRIFGPKETILAI 443
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 IVNGTPVNATEFVWKANTGWYVAFRIKTPKEGTVTTPYCAFISSVANISSLPNY-----277
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 444 ILLAIFLQ--DIYLYIIGFEIIL-----NTVIIP-----NLSLPSYLSPH 484
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 -----TELYLED 284
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 FNHQHYSLESPEEQIFLED 504
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 15

T09620

probable lectin 2 precursor - alfalfa

C:Species: Medicago sativa (alfalfa)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000

C:Accession: T09620

R:Brill, L.M.; Pietermel, V.R.

Submitted to the EMBL Data Library, March 1998

A:Description: Legume seed lectin genes: sequence of Molec2 from Alfalfa, Alfalfa and Sw

A:Reference number: Z16780

A:Accession: T09620

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-279 &lt;BRI&gt;

A:Cross-references: EMBL:Y16754

A:Experimental source: Cultivar Chief

C:Genetics:

A:Gene: lec2

C:Superfamily: plant lectin

C:Keywords: calcium; glycoprotein; lectin

F:1-25/Domain: signal sequence #status predicted &lt;SIG&gt;

F:26-279/Product: probable lectin 2 #status predicted &lt;MAT&gt;

Query Match 5.7%; Score 97.5; DB 2; Length 279;

Best Local Similarity 21.5%; Pred. No. 3.1; Indels 75; Gaps 17;

Matches 67; Conservative 53; Mismatches 116;

QY 1 MSKKKPVIVSILTILLVOAIYFEVKYHTSEDKSTNTSSTPPQTTL-----STTKVLKI 54

Db 1 MSSSNPFS--ILSISILTFILLMNKVSNAETTSFISITKFPDQKNLIFQGDAAETTTGKL 58

QY 55 RYPDDGEWPGAPIDKDGDNPGFYEINLWNLNATG-----FAEMTYNLT 101

Db 59 KL-----TXAVKNSIGRALYSAPIHWD--SKTGSVANPETFTTTITAPNTYNVAD 108

QY 102 GVLHYVQOLDNIVLRDRSNWVHGYPFIFGYNKPNANYATDGPILPSPKVSNTLDFYLT 161

Db 109 GLAFFIAPIDT---QPKSNSQGYLGVP-DSKTYQESIQT-----VAVEIDTFYV 155

QY 162 SYKLEPKN-----GLPINFAESWLTREAWRTGINSDEQEVMIWIIYDGLQPAAGSK 213

Db 156 DWDPKPGNTSSTGRHIGINVN-SIXS-ITTVPSWL--VNNKKANVVI-----GFNGATNV 206

QY 214 VK-EIIVVPIVNGT-----PVNATEFVW-----KANIGWEYVAFRIKTPKEGTVTIPY 262

Db 207 LSVDEVYPLVRHYTLTSHVPLKDVVPEWVRIGFSAAATGDEYAEHDIFSWSPDKNLNL---- 263

QY 263 AFISVAANISS 273

Db 264 GPDNINAVSS 274

Search completed: June 29, 2004, 20:13:54

Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 19:55:03 ; Search time 17 Seconds  
(without alignments)  
977.081 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701

Sequence: 1 MSKKKFFVSVILTLLVQAI.....HLEWWTITNLTPLDRPLIS 319

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112.5	6.6	264	1	GUNS_ERWCA
2	103	6.1	421	1	D3ML_MOUSE
3	99.5	5.8	597	1	PEPX_BACAA
4	99	5.8	237	1	GUN_ASPAC
5	99	5.8	1680	1	COS_MOUSE
6	97.5	5.7	710	1	PRCE_MOUSE
7	95.5	5.6	2316	1	FPFZ_RAT
8	94.5	5.6	247	1	FLAL_THEVO
9	94	5.5	772	1	CIPB_CLOTH
10	93	5.5	608	1	YD93_METJA
11	93	5.5	785	1	VP35_SCHPO
12	92	5.4	417	1	IF_MOUSE
13	91	5.3	1853	1	CIFP_CLOTH
14	90.5	5.3	1481	1	APU_THEST
15	90	5.3	463	1	GUN_BACSP
16	89.5	5.3	710	1	PFCE_PIG
17	89.5	5.3	2352	1	MOKC_SCHPO
18	89	5.2	703	1	CDGT_BACS2
19	88	5.2	1183	1	CNA_STAAR
20	87.5	5.1	593	1	CSG_METFE
21	87.5	5.1	1848	1	CBPA_CLOCL
22	87	5.1	312	1	YC90_METJA
23	87	5.1	485	1	ENT_ENTCO
24	87	5.1	3672	1	LML2_CABEL
25	86.5	5.1	710	1	PPCE_BOVIN
26	86.5	5.1	783	1	TRF2_THEVO
27	86	5.1	239	1	GUNA_ASPKA
28	86	5.1	879	1	MANB_CAPHI
29	86	5.1	1256	1	ATL_STAAR
30	86	5.1	1664	1	RPAL_YEAST
31	85.5	5.0	626	1	GPBA_HUMAN
32	85.5	5.0	666	1	NEPI_THEVU
33	85.5	5.0	839	1	YDDE_HABIN

## RESULT 1

ID	GUNS_ERWCA	STANDARD;	PRT;	264 AA.
AC	PI6630;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Endoglucanase S precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase S)			
DE	(Cellulase S).			
GN	CELS.			
OS	Erwinia carotovora.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pectobacterium.			
OX	NCBI_TaxID:554;			
[1]				
RN	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RP	STRAIN=SCC3193;			
RC	MEDLINE=90337352; PubMed=2379837;			
RX	Saarilahti H.T., Henrissat B., Palva E.T.;			
RA	"Cels: a novel endoglucanase identified from Erwinia carotovora			
RT	subsp. carotovora";			
RL	Gene 9039-14(1990);			
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.			
CC	-!- SIMILARITY: BELONGS TO CELLULASE FAMILY H (FAMILY 12 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	M32399; AA24817.1; -			
DR	PIR; J00328; J00328.			
DR	InterPro; IPR008985; ConA-like lec.gl.			
DR	InterPro; IPR002594; Glyco_hydro_12.			
DR	Pfam; PF01670; Glyco_hydro_12; 1			
DR	ProDom; PD004316; Glyco_hydro_12; 1.			
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
FT	SIGNAL			
FT	CHAIN 1 32			
FT	SEQUENCE 264 AA; 29757 MW; E6D61388950C77AA CRC64;			
SQ				
Query Match	6.6%; Score 112.5; DB 1; Length 264;			
Best Local Similarity	21.8%; Pred No. 0.13; Mismatches 69; Gaps 14;			
Matches	55; Conservative 38; Indels 90; Indels 69; Gaps 14;			
QY	68 DKDGGNPEFYIEINLWILNATGPAENTYNTSGVLHYVQQLDNIVLDRSNW----- 121			
DB	40 DKLYFGNNKYLFNNVWGKDEIKGWOQIFY-----NSPISMGNNHWPST 86			
QY	122 --VHGYPBIFVGNKPNANYATDGPILPSPKSVNLTDFYLTISYKLEPKNGLPINFAIES 179			

87 HSKAYPSLVSG-WHWTAGYTENGLPI--QLSSNKITSNTVYSIKATG--TYNAAYDI 141  
190 WL---TREAWTTGINSDEQWMTWYDGLQAGSKVKIIVPIVNGTFFVNAATEVWK 236  
142 WFHTTKANWD-----SSPTDEIMTWJNDTNAAGPDAGYIETVFLG-----DSSMNVFK 189  
237 ANI-----GWYVAFRIKTKIKGTIVTIPYGAISVAANISSLPNY-----TELY 281  
190 GWINADNGGWNVSF-VHT-----NSASLNIRHFTDYLVTQKQWMSDEKY 237  
282 LEDVEIGTE-FG 292  
238 ISSVEFGTEIFG 249  
RESULT 2  
ID DM3L MOUSE STANDARD; PRT; 421 AA.  
AC Q9CWR8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE DNA (cytosine-5)-methyltransferase 3-like.  
GN DNMT3L.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=11306809;  
RX MEDLINE=21203556; PubMed=11306809;  
RA Apolita U., Lyle R., Krohn K., Antonarakis S.E., Peterson P.;  
RT "Isolation and initial characterization of the mouse Dnmt3l gene.";  
RL Cytogenet. Cell Genet. 92:122-126(2001).  
[2]  
RN SEQUENCE FROM N.A.  
RA Shaoping X., Hata K., Li E.;  
RT "Full-length cDNA of a murine Dnmt3-like gene.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;  
RC MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
[4]  
RN FUNCTION.  
RP MEDLINE=21625316; PubMed=11719592;  
RX Bourc'his D., Xu G.L., Lin C.S., Bollman B., Bestor T.H.;  
RA "Dnmt3l and the establishment of maternal genomic imprints.";  
RT Science 294:2536-2539(2001).  
RL CC -!- FUNCTION: Probably not catalytically active as it has lost the  
CC active site residues. May function not directly as a DNA  
CC methyltransferase but as a regulator of methylation at imprinted  
CC loci. It is required specifically for the establishment of genomic

imprints but is dispensable for their propagation. It is essential  
for the de novo methylation of single-copy DNA sequences.  
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
-!- TISSUE SPECIFICITY: Expressed in testis, thymus, ovary, and heart.  
-!- SIMILARITY: Belongs to the C5-methyltransferase family.  
-!- SIMILARITY: Contains 1 ADD-type zinc finger.  
This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL: AJ404467; CAB94726.1; -  
DR EMBL: AF220524; AAF73868.1; -  
DR EMBL: AK010434; BAB26936.1; -  
DR MGI: 1859287; Dnmt3l.  
GO: GO:0005737; Cytoplasm; IDA.  
GO: GO:0005720; C:nuclear heterochromatin; IDA.  
GO: GO:0005634; C:nucleus; ISS.  
GO: GO:0019899; F:enzyme binding; ISS.  
GO: GO:0016564; F:transcriptional repressor activity; ISS.  
GO: GO:0006306; P:DNA methylation; ISS.  
GO: GO:0001701; P:embryonic development (sensu Mammalia); IMP.  
GO: GO:0006349; P:imprinting; ISS.  
GO: GO:0007283; P:spermatogenesis; ISS.  
KW Zinc-finger; Zinc; Metal-binding; Nuclear protein.  
FT ZN FING 87 179 ADD-TYPE.  
SQ SEQUENCE 421 AA; 47992 MW; 6C996D220C6F6D83 CRC64;  
Query Match 6.1%; Score 103; DB 1; Length 421;  
Best Local Similarity 20.6%; Pred. No. 1.3;  
Matches 32; Conservative 22; Mismatches 53; Indels 48; Gaps 5;  
QY 68 DXDGDGNPEFYIEINLW-----NIINATGFAEMTYNLTSGVLHYVQQLDN 112  
DB 203 DQEGAGPMIYKTVSAWKQPVRLSLPRNIDKVLKSLGFLBSGSGSGGTLYVEDVTN 262  
QY 113 IVLRDRSNVHGYPEIFVGNKPNANYATDGPILPSKVSNLTDVLIISYKLPKNGLP 172  
DB 263 VVRDVEKW--GPFDLVYGSTQ-----PLGSSCDRCPCGWINQFHRI----- 302  
QY 173 INPAIESMLTREATRTTGINSDEQWMTWYDGL 207  
DB 303 LQVALP-----RQSSQRPFFWIFMDNL 324  
RESULT 3  
ID PEPX\_BACAA STANDARD; PRT; 597 AA.  
AC Q81PE9;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative Xaa-Pro dipeptidyl-peptidase (EC 3.4.14.11) (X-Pro  
DE dipeptidyl-peptidase) (X-prolyl-dipeptidyl aminopeptidase) (X-PDAP).  
GN BA2850.  
OS Bacillus anthracis (strain Ames).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=198094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608414; PubMed=12721629;  
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
RA Kolonay J.F., Seaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
RA DeBoy R.F., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,



OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. PubMed=2303408;  
 RX MEDLINE=90153853; PubMed=2303408;  
 RA Wetzel R.A., Fleischer D.T., Haviland D.L.;  
 RT "Deficiency of the murine fifth complement component (C5). A 2-base  
 pair gene deletion in a 5'-exon.";  
 RL J. Biol. Chem. 265:2435-2440(1990).  
 RN [2]  
 RP SEQUENCE OF 41-1680 FROM N.A.  
 RX MEDLINE=87185363; PubMed=2436653;  
 RA Wetzel R.A., Ogata R.T., Tack B.F.;  
 RT "Primary structure of the fifth component of murine complement.";  
 RL Biochemistry 26:737-743(1987).  
 CC -!- FUNCTION: Activation of C5 by a C5 convertase initiates the  
 CC spontaneous assembly of the late complement components, C5-C9,  
 CC into the membrane attack complex. C5b has a transient binding site  
 CC for C6. The C5b-C6 complex is the foundation upon which the lytic  
 CC complex is assembled.  
 CC -!- FUNCTION: Derived from proteolytic degradation of complement C5,  
 CC C5 anaphylatoxin is a mediator of local inflammatory process. It  
 CC induces the contraction of smooth muscle, increases vascular  
 CC permeability and causes histamine release from mast cells and  
 CC basophilic leukocytes. C5a also stimulates the locomotion of  
 CC polymorphonuclear leukocytes (chemotaxis) and direct their  
 CC migration toward sites of inflammation (chemotaxis).  
 CC -!- SUBUNIT: C5 precursor is first processed by the removal of 4 basic  
 CC residues, forming two chains, beta and alpha, linked by a  
 CC disulfide bond. C5 convertase activates C5 by cleaving the alpha  
 CC chain, releasing C5a anaphylatoxin and generating C5b (beta chain  
 CC + alpha' chain).  
 CC -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.  
 CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.  
 CC -!- SIMILARITY: Contains 1 NTR domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M35525; AAA37349.1; -;  
 DR EMBL; M35526; AAA37348.1; -;  
 DR PIR; A35530; C5MS.  
 DR HSP; P01031; 1KJS.  
 DR MGD; MGI:96031; Hc.  
 DR InterPro; IPR002890; A2M N.  
 DR InterPro; IPR009048; AM\_receptor\_bind.  
 DR InterPro; IPR000020; Anaphylatoxin.  
 DR InterPro; IPR001840; Anaphylatoxin.  
 DR InterPro; IPR001599; MacroglobulinA2.  
 DR InterPro; IPR001134; Netrin C.  
 DR InterPro; IPR008930; Terp\_cyc\_toroid.  
 DR Pfam; PF00207; A2M; 1.  
 DR Pfam; PF01835; A2M N; 1.  
 DR Pfam; PF01821; ANA7O; 1.  
 DR Pfam; PF01759; NTR; 1.  
 DR PRINTS; PRO0004; ANAPHYLATOXN.  
 DR ProDom; PD003264; Anaphylatoxin; 1.  
 DR SMART; SM00104; ANATO; 1.  
 DR SMART; SM00643; C345C; 1.  
 DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; FALSE\_NEG.  
 DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
 DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
 DR PROSITE; PS01189; NTR; 1.  
 DR PROSITE; PS01189; NTR; 1.  
 KW Complement pathway; Complement alternate pathway; Glycoprotein;  
 KW Plasma; Membrane attack complex; Cytolysis; Inflammatory response;  
 KW Signal.

FT SIGNAL 1 18  
 FT CHAIN 19 1680  
 FT CHAIN 19 674  
 FT PROPEP 675 678  
 FT CHAIN 679 1680  
 FT PEPTIDE 679 755  
 FT CHAIN 756 1680  
 FT DOMAIN 702 736  
 FT DOMAIN 1536 1679  
 FT DISULFID 702 728  
 FT DISULFID 703 735  
 FT DISULFID 715 736  
 FT CARBOHYD 427 427  
 FT CARBOHYD 915 915  
 FT CARBOHYD 1119 1119  
 FT CARBOHYD 1633 1633  
 FT VARIANT 216 216  
 FT VARIANT 217 1680  
 FT SEQUENCE 1680 AA; 188877 MW; 81E5A16FAC7D95C CRC64;  
 SQ  
 Query Match 5.8%; Score 99; DB 1; Length 1680;  
 Best Local Similarity 22.8%; Pred. No. 15;  
 Matches 66; Conservative 32; Mismatches 95; Indels 96; Gaps 13;  
 QY 81 INLWNI-----LNATGFAEMTYNLTSGVLYHYVQOLDNIVLRDRSNWVHGYEIPYG--- 131  
 DB 1 MGLWGLCELLFLDKTWGQEQYIVISAPKILRVGSSENVVIO-----VHGYTEAFDATLS 55  
 QY 132 -----NKPNNVATDGPPLPSKYSNLTDFYL-TISYKLEPKN 169  
 DB 56 LKSYDPKKVTFSSGYVNSPENKFNQAAALLTQPNQVRESPPVSHVLEVVSKHFSKSK 115  
 QY 170 GLPINPAIESLWLTREAWRTGI-----NSDQEVMIWY--DGLQFA----- 210  
 DB 116 KIPITY-----NNGILFIHTDKPVYTPQSKIRVYSLGDDLKPAKRETVLTF 163  
 QY 211 ----GSKVKEIVVPIVING-----TPVNATFEW--KANIG-----WEY 243  
 DB 164 IDPEGSEV-DIVSENDYTGIIISFPDFKIPSNPKYGVWTKANKYKDKFTTGTAYFEIKEY 222  
 QY 244 VAFRIKTPKEGTIPYGAFTSVAANTSSLPNTLYLEDVEIGTEFG 292  
 DB 223 VLPFRSVSIELERTFIGNKFNKFTITVKARYFNKV-VFDAEYVAFG 270  
 RESULT 6  
 ID PPCE\_MOUSE STANDARD; PRT; 710 AA.  
 AC QOQUR6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Prolyl endopeptidase (EC 3.4.21.26) (Post-proline cleaving enzyme)  
 DE (PE).  
 GN PREP OR PEP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98207037; PubMed=9538240;  
 RA Ishino T., Ohtsuki S., Homma K., Natori S.;  
 RT "cDNA cloning of mouse prolyl endopeptidase and its involvement in DNA  
 RT synthesis by swiss 3T3 cells.";  
 RL J. Biochem. 123:540-545(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99377037; PubMed=10446174;  
 RA Kimura A., Yoshida I., Takegi N., Takahashi T.;  
 RT "Structure and localization of the mouse prolyl oligopeptidase gene.";  
 RL J. Biol. Chem. 274:24047-24053(1999).

[3] SEQUENCE FROM N.A.  
 RP STRAIN=FVE/N; TISSUE=Limb, and Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Wozny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzley D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Cleaves peptide bonds on the C-terminal side of prolyl  
 CC residues within peptides that are up to approximately 30 amino  
 CC acids long.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of Pro-|-Xaa >> Ala-|-Xaa in  
 CC oligopeptides.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILILARY: Belongs to peptidase family S9A.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB007631; BAA8239.1; -  
 CC EMBL; AB022053; BAA83071.1; -  
 CC EMBL; AB022047; BAA83071.1; JOINED.  
 CC EMBL; AB022048; BAA83071.1; JOINED.  
 CC EMBL; AB022049; BAA83071.1; JOINED.  
 CC EMBL; AB022050; BAA83071.1; JOINED.  
 CC EMBL; AB022051; BAA83071.1; JOINED.  
 CC EMBL; AB022052; BAA83071.1; JOINED.  
 CC EMBL; BC012869; AAH12869.1; -  
 CC EMBL; BC050830; AAH50830.2; -  
 CC PIR; JWC080; JWC080.  
 CC HSP; P23687; 10FS.  
 CC MEROPS; S09.001; -  
 CC MGI; MGI:1270863; Prep.  
 CC GO; GO:0005737; Cytoplasm; IDA.  
 CC GO; GO:0005634; C:nucleus; IDA.  
 CC InterPro; IPR002471; Pept\_S9\_AS.  
 CC InterPro; IPR001375; Peptidase\_S9.  
 CC InterPro; IPR002470; Peptidase\_S9A.  
 CC InterPro; IPR004106; Peptidase\_S9A\_N.  
 CC InterPro; IPR000379; Ser-estr.  
 CC Pfam; PF00326; Peptidase\_S9; 1.  
 CC Pfam; PF02897; Peptidase\_S9\_N; 1.  
 CC PRINTS; PR00862; PROLIPOPTASE.  
 CC PROSITE; PS00708; PRO-ENDOPEP\_SER; 1.  
 CC Hydrolase; Serine protease.  
 KW ACT\_SITE 554 554 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 641 641 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 680 680 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 710 AA; 80751 MW; 13010D5DCAV3C03 CRC64;

Query Match

Best Local Similarity 5.7%; Score 97.5; DB 1; Length 710;

Pred. No. 6.5;

Matches 69; Conservative 52; Mismatches 130; Indels 95; Gaps 16;  
 QY 22 FVEKHTSDKSTNTSTSTPQTTL-----STTKVLKIRYPDDGEW-PCAPIDKQGD 72  
 DB 186 FYNFPOQDKSGDGTETSTNHLQKLCYHVLGTQDSEDILCAEFDPDEPKWGGAEISDDGR 245  
 QY 73 GNPEFYIENLW---NILNATGAEMTY--NUTSGVLHYVQOLDNIVLRDNRNWHGVPE 127  
 DB 246 ----YVLLSIWEGCDPVNRLMYCLQQEPNGITGILKWKLIDN--FEGEYDYVTNEGT 298  
 QY 128 IFYGNKPWNA-NVA---TDGPILPSK-----VSNLTFYLTISKLEBPK 168  
 DB 299 VTFKTRNSPNRNLINDFTDDESKKVLPEHEKDVLEWVAVCURSNFLVLCYLHVK 358  
 QY 169 NGLPINFAIESWLTR----EAWRTTGINSDEQVMIWYDGLQAG-----SKVKEIVV 219  
 DB 359 NILQLHDLTGTALLTFPLDVGVSGRKKDSIFYQTSFLSPGVYIHCDLTKEELE 418  
 QY 220 PIIVNGTPVNATFEVWKANIGWEYVAFRIKTKIKEGTVIP----- 260  
 DB 419 PMWFREVTY-----KGIDAADYQTIQIFYPSKQGT-KIPMFIVHKGKIKLDGSHAPL 470  
 QY 261 --YGAFISVAANISSLPNT-----ELYLEDVIEIGTEFG 292  
 DB 471 YGVGGF-----NISITPNYSVSLIFVRHMGGLAVANIRGGGEYV 511  
 RESULT 7  
 PTPZ\_RAT  
 ID PTPZ RAT STANDARD; PRT; 2316 AA.  
 AC Q62656; Q62621;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Receptor-type protein-tyrosine phosphatase zeta precursor  
 DE (BC 3.1.3.48) (R-PTP-zeta) (Phosphacan) (3F8 chondroitin sulfate  
 DE proteoglycan) (3H1 keratan sulfate proteoglycan).  
 DE PTPRZ1 OR PTPRZ OR PTPZ.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=96063026; PubMed=7579589;  
 RA Maurel P., Meyer-Puttlitz B., Flad M., Margolis R.U.,  
 RA Margolis R.K.;  
 RT "Nucleotide sequence and molecular variants of rat receptor-type  
 RT protein tyrosine phosphatase-zeta/beta.";  
 RL DNA Seq. 5:323-328(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND PARTIAL SEQUENCE.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=94195772; PubMed=7511813;  
 RA Maurel P., Rauch U., Flad M., Margolis R.K., Margolis R.U.;  
 RT "Phosphacan, a chondroitin sulfate proteoglycan of brain that  
 RT interacts with neurons and neural cell-adhesion molecules, is an  
 RT extracellular variant of a receptor-type protein tyrosine  
 RT phosphatase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:2512-2516(1994).  
 RN [3]  
 RP BINDING TO N-CAM AND NG-CAM (PHOSPHACAN).  
 RX MEDLINE=95096181; PubMed=7528221;  
 RA Milev P., Friedlander D.R., Sakurai T., Karthikeyan L., Flad M.,  
 RA Margolis R.K., Grunet M., Margolis R.U.;  
 RT "Interactions of the chondroitin sulfate proteoglycan phosphacan, the  
 RT extracellular domain of a receptor-type protein tyrosine phosphatase,  
 RT with neurons, glia, and neural cell adhesion molecules.";  
 RL J. Cell Biol. 127:1703-1715(1994).  
 RN [4]  
 RP BINDING TO TENASCIN (PHOSPHACAN).  
 RX MEDLINE=94216329; PubMed=7512960;







```
RESULT 8
FLAI THEVO
ID FLAI THEVO STANDARD; PRT; 247 AA.
AC P57719;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Flagellin B1 precursor (41 kDa flagellin).
GN TV0607 OR TV0598631.
OS Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
RN [2]
RP SEQUENCE OF 21-40.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=96146545; PubMed=8550330;
RA Faguy D.M., Bayley D.P., Kostyukova A.S., Thomas N.A., Jarrell K.F.;
RT "Isolation and characterization of flagella and flagellin proteins
RT from the Thermoacidophilic archaea Thermoplasma volcanium and
RT Sulfolobus shibatae."
RL J. Bacteriol. 178:902-905(1996).
RC J.
CC -!- FUNCTION: Flagellin is the subunit protein which polymerizes to
CC form the filaments of archaeal flagella.
CC -!- PFM: Glycosylated.
CC -!- SIMILARITY: Belongs to the archaeal flagellin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000992; BAB59749.1;
DR InterPro; IPR002774; Arch_flagellin.
DR Pfam; PF01917; Arch_flagellin; 1.
KW Flagellum; Glycoprotein; Complete proteome.
FT PROPEP 1 20
FT CHAIN 21 247 FLAGELLIN B1.
FT CONFLICT 21 22 ET -> GI (IN REF. 2).
SQ SEQUENCE 247 AA; 25729 MW; 63CD4B3024C3E22 CRC64;
Query Match 5.6%; Score 94.5; DB 1; Length 247;
Best Local Similarity 25.8%; Pred. No. 2.9;
Matches 47; Conservative 28; Mismatches 68; Indels 39; Gaps 9;
QY 7 VIVSILTLVLAQVIFVEKYHTSDKSTNSSTPQTT-----LSTTKVLKIRY 56
DB 29 VFIAVLVAANAATVLTAGTGLQKATSTGCTQVSTGICQVNSIYGLDSKSV---- 84
QY 57 PDDG--EWPGAPIDKDGNGNPEFYIEINLWNILNATGFAEMTYNLTGVLHYVQQLDNIV 114
DB 85 PTHGVIEMLAQISITAGSSP-----INLANVT-----ISLTHGVGSASLTIV-V-GLENIG 133
QY 115 LRDRSNWVHGYPFIYFG-NKPWANY-----ATDGP-----IPLPKVSNLTDFTYLTIS 162
DB 134 NATVINDVYGENSAVGGTNNVFNSSYFKTINGASNGSKHFAILLVSDFTNSMTAQYPVIS 193
QY 163 YK 164
DB 194 YE 195
RESULT 9
CIPB CLOTM
ID CIPB CLOTM STANDARD; PRT; 772 AA.
AC Q01866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein S1/SL)
DE Cellulose integrating protein B (Fragment).
GN CIPB
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS;
RX MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
RA Gilbert H.J.;
RT "Identification of the cellulose-binding domain of the cellulosome
RT subunit S1 from Clostridium thermocellum YS."
RL FEMS Microbiol. Lett. 78:181-186(1992).
CC -!- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
CC PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -!- SUBCELLULAR LOCATION: Cell surface.
CC -!- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY
CC THE CATALYTIC COMPONENTS OF THE CELLULOSE.
CC -!- SIMILARITY: Contains at least 3 cohesin domains.
CC -!- SIMILARITY: Contains 2 dockerin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X68233; CAA48312.1;
DR HSPSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008985; Cohesin.
DR InterPro; IPR002102; Cohesin.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00942; CEM_3; 1.
DR Pfam; PF00963; Cohesin; 3.
DR Pfam; PF00404; Dockerin_1; 2.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN <1 80 COHESIN 1.
FT DOMAIN 81 93 LINKER (PRO/THR-RICH).
FT DOMAIN 94 240 COHESIN 2.
FT DOMAIN 241 272 LINKER (PRO/THR-RICH).
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 LINKER (PRO/THR-RICH).
FT DOMAIN 462 607 COHESIN 3.
FT DOMAIN 710 733 DOCKERIN 1.
FT DOMAIN 743 766 DOCKERIN 2.
SQ SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;
Query Match 5.5%; Score 94; DB 1; Length 772;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 66; Conservative 30; Mismatches 127; Indels 60; Gaps 13;
```

QY 34 TSNTSSTPQTTLTKVLRYPDDGWPAGAPDKDGNPFPVIEINLWNLNATGPA 93  
 DB 265 TINTPTANTPVSNG--LKVEF-----YNNPSTDTNSINPQKV-----TNTG 307  
 QY 94 EMTYNLTSGVLHYVQOQDNLVLRDRSMVHGYPFIFGKGNKPNWA-NYATDGPILPSPKVS 152  
 DB 308 SSAIDLSKLTLLRYVYVTDG--QKDTFWC-DHAAIIGNSGYNGITSNVKGTFFVKMSST 364  
 QY 153 NLTDYFLISY---KLEPKNGLPI--NFAIESLTLREARWTTGINSDEQ----- 196  
 DB 365 NNADTYLEISFTGGTLEGAHVQIGRFKANDWSNYTOSNDYSFKSRQSFVENDQVTAYL 424  
 QY 197 -EVMIIWYDGLQAGSKVEITVPIIVNGTFPVNATFEVWKANIGWYVAFRIKTIKEG 255  
 DB 425 NGVLWV---GKEPGGS-----VVPSTQVTPPTATTTPPSDDPNAIK--IKVD 473  
 QY 256 TVTIYPGAFISVANISLSP-----NYTELYLEDVBI 287  
 DB 474 TVNAKPGDTVNIIPVRFSGIPSGKIANCDFVYSPDNVLEIIEI 516

RESULT 10  
 YD93\_METJA  
 ID YD93\_METJA STANDARD; PRT; 608 AA.  
 AC Q58788;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ1393.  
 GN MJ1393.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts C.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -!- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U67579; AAB99403.1; -.  
 CC F1R; H64473; H64473.  
 CC TIGR; MJ1393; -.  
 CC KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 4 24 POTENTIAL.  
 SQ SEQUENCE 608 AA; 56768 MW; 010FAFLC29F8C73C CRC64;  
 Query Match 5.5%; Score 93; DB 1; Length 608;  
 Best Local Similarity 20.1%; Pred. No. 12;  
 Matches 82; Conservative 54; Mismatches 128; Indels 144; Gaps 22;  
 QY 4 KKFVIVSIL-TILLVQAIY-----FVEKYHTSEDKSTNSSTPQTTLSTTKV 51

DB 2 RKLIFMALLMSLLFIGTVEGYGNGPLVYAYVEKYN-----TGNTT----- 43  
 QY 52 LKIRYPDDGWPAGAPDKDGNPFPFIE-INLWNLNATGPAEMTYNLTSGVLHYVOOL 110  
 DB 44 -----GDLVSTIESITGYVINNTG--TTINDT---LYDVWA 78  
 QY 111 DNIVLRDRSNWVGHPFIEF-----YGNKPWNAN----- 138  
 DB 79 VNI-----SNNTGT-PEVYVNGTPKGVFTSESSAPATNLP-NANTYIHIPILPNNSYVII 131  
 QY 139 -YATDGP-----PLSKVSNLTFDYLTISYKLE--PKNGLPINFAIESWL 181  
 DB 132 KFAIDKSIITGVPLIINETVSDTKIISERLSNWSVINSRVNSALPATDTPVSVIMTKYL 191  
 QY 182 TR-----EAW-----RTTGINSDEQEVMIWI-----YYDGLQFAG-----SKYEIVV 219  
 DB 192 SNDPNNGYSDTWNFLNITGAIANEGSITLWDGPFYLPFGYNDLSLTWTVGVVINTKNTATITI 251  
 QY 220 PIIVNGTFPVNATFEVWK---ANIGWEYVAFRIKTIKEGTVPIYGAFISVAAN--ISSL 274  
 DB 252 NITGNNTYTRTGTLMKYGFAVIFEFNGTKSGTKI-EGIYATGCGVSATKEGPFNLAS 310  
 QY 275 PNYTELYLEDVEITGTEFTPTSTSAHLEWITN-----ITLPLDRPLI 318  
 DB 311 SGKYEIWEESANVSKASSYVFNLTHTVIAVNGSNPVLDPFNITLL 358

RESULT 11  
 VP35\_SCHPO  
 ID VP35\_SCHPO STANDARD; PRT; 785 AA.  
 AC Q74552; P78830;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vacuolar protein sorting-associated protein vps35.  
 GN VPS35 OR SPC777.13.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown N., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders R., Squares S., Stevens K.,  
 RA Skelton J., Simmonds M., Squares R., Tavey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Vackaert G., Aert R., Robben J., Schaefer M., Mueller-Auer S.,  
 RA Weltjens I., Vanstreels B., Rieger M., Schaefer M., Hilbert H.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
 RA Sipakovsky G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 RN [2]  
 RP SEQUENCE OF 279-785 FROM N.A.

RC STRAIN=PR745;  
 RX MEDLINE=98162722; PubMed=9501991;  
 RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in Schizosaccharomyces pombe  
 cDNAs.";  
 RL DNA Res. 4:363-369(1997).  
 CC -!- FUNCTION: May play a role in vesicular protein sorting, analogous  
 CC to the yeast retromer proteins (By similarity).  
 CC -!- SUBUNIT: Probably a component of a membrane-associated  
 CC multiprotein complex (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Associated with an intracellular membrane  
 CC (By similarity).  
 CC -!- SIMILARITY: Belongs to the Vps35 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AL031532; CAA20717.1; -;  
 DR EMBL; D89178; BAAL3840.1; -;  
 DR PIR; I11719; T11719.  
 DR GeneDB SPombe; SPCC77.13; -;  
 DR InterPro; IPR005378; Vps35.  
 DR Pfam; PF03635; Vps35; 1.  
 KW Transport; Protein transport; Membrane.  
 FT CONFLICT 337 337  
 FT SEQUENCE 785 AA; 90608 MW; 0BA4BLCFBF58D68 CRC64;  
 SQ  
 Query Match 5.5%; Score 93; DB 1; Length 785;  
 Best Local Similarity 20.7%; Pred. No. 16;  
 Matches 75; Conservative 56; Mismatches 117; Indels 114; Gaps 19;  
 QY 8 IVSILTLLVQAIFYVEKYHSEDKSTNTSSTPQTTLSTTKV-LKIRYPDGEWGP 66  
 Db 244 VTQLVAMNLTDYVQREYES-DSSNEDESEVTE-KLGDIKINEVQKDECECGDK 301  
 QY 67 IDKDGNGNPEFYIEINLWN-----ILNATGP-AEMTYNLTSGLVHYVQQLDNIVLRDSN 120  
 Db 302 VIP-----PEVAIQELVSHVHVIEQSRGLPLDCIVLSILSIL-----NPLR---- 345  
 QY 121 WHGYPEIFYGNKPNAN-----YATDGPILPDKVSNLTD-----FYLT 160  
 Db 346 -----CYPKQYADRVFPQYNEHIINQPSLSALHERPQKSLCAILLPLTYFPS 397  
 QY 161 ISYKLEPKNGLPINFAIESWLTREAWR-----  
 Db 398 PSYCLELQNFPLPVNAQDPNLRVDIARMTVQKITEKGHSISELTAQELLGFVSVITEKK 457  
 QY 190 GINS--DEQYMIWIIY-----DGLQAGSKYKEIWPFIIVNGTPVNA 230  
 Db 458 GVDLSDDIQNALVMVHLYNDDPOIQIBILRSKLDTFKAGENVK-YLLFPVWNRCIFLA 516  
 QY 231 -TEEV-----W--KANTGWVAVFIKTPKEG-TVPIPGAFISVAANSSLPNVTLY 281  
 Db 517 RNFRIKCMDAEKVRLLWEFVNTCINLVYKNGSDSLCLALYLUS-AAEMADQENYDFPA 575  
 QY 282 LE 283  
 Db 576 YE 577  
 RESULT 12  
 IF MOUSE  
 ID IF MOUSE  
 AC P52787;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Intrinsic factor precursor (IF) (INF) (Gastric intrinsic factor).

GN GIF.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Stomach;  
 RX MEDLINE=94075347; PubMed=8253786;  
 RA Lorenz R.G., Gordon J.I.;  
 RT "Use of transgenic mice to study regulation of gene expression in the  
 RT parietal cell lineage of gastric units.";  
 RL J. Biol. Chem. 268:26559-26570(1993).  
 CC -!- FUNCTION: Promotes absorption of the essential vitamin cobalamin  
 CC (Cbl) in the ileum by specific receptor-mediated endocytosis.  
 CC -!- TISSUE SPECIFICITY: Gastric mucosa.  
 CC -!- SIMILARITY: Belongs to the eukaryotic cobalamin transport proteins  
 CC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; L24191; AAA37802.1; -;  
 DR EMBL; L24192; AAA37881.1; -;  
 DR PIR; A49684; A49684.  
 DR MGD; MGI:1202394; Gif.  
 DR InterPro; IPR002157; Cobalamin\_bind.  
 DR Pfam; PF01122; Cobalamin\_bind; 1.  
 DR PROSITE; PS00468; COBALAMIN\_BINDING; 1.  
 KW Transport; Cobalt transport; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 417  
 FT DISULFID 26 246  
 FT DISULFID 103 288  
 FT DISULFID 143 182  
 FT CARBOHYD 86 86  
 FT CARBOHYD 311 311  
 FT CARBOHYD 330 330  
 FT CARBOHYD 413 413  
 FT SEQUENCE 417 AA; 45395 MW; 421713BAE6358DC2 CRC64;  
 SQ  
 Query Match 5.4%; Score 92; DB 1; Length 417;  
 Best Local Similarity 21.8%; Pred. No. 8.8;  
 Matches 78; Conservative 41; Mismatches 133; Indels 106; Gaps 19;  
 QY 12 LITLLVQAIFYVEKYHSEDKSTNTSSTPQTTLSTTKVLRYPDGEWGPIDK 69  
 Db 4 LTYLLSLVWAVAGTSTRAQSS-----CSVPPQQPVDGLQALMENSVTDSDFP----- 53  
 QY 70 DGDGPEFYIEINLWNILNATGAEMTY-----NLTSGLVHY-VQQLD----- 111  
 Db 54 ----NFSILIANLAGAVNEAQLTYQLMASDSANLTSGQLALTVMALTSRCDPSGK 109  
 QY 112 -NIVLRDRNWHGYP-----EIPYG-----NKPWANYATD--GPIPLP 148  
 Db 110 VSTLLKKMENWSPSPGAESEAFYGFGLAILALCOCKSEATLPIAVAFKTLMEFPSPN 169  
 QY 149 SKVSNLTDFYLTISYKLEPKNGLPINFAIESWLTREAWR-----TGINSDEQVMIWIIYD 205  
 Db 170 VDTGAVATLALTCMY-----NKIPYG-----SQENYEDLFGQALKAVEIKSLIKAD 217  
 QY 206 G-----LQPAGSKYKEIWPFIIVNGTPVNAFVWKNIGWVAVPRIKTPKEGTVTIPY 261  
 Db 218 GIIGDIYSTGLAMQALSV-----TPEQPT-KKWDG-----EKTMTILNEIKQKGFQNP- 265  
 QY 262 GAFISVAANISSLPNVTLYLEDVIG-----TEGTPSTTSIAHLEWITNITL 310  
 Db 266 ----MSIAQILPSLKGKTYLDVFPQVTCGPDHEVPPLTDYPTVPVPS-----VSNITV 314



```

FT STRAND 467 468
FT TURN 470 471
FT STRAND 473 482
FT TURN 483 484
FT STRAND 488 489
FT TURN 491 492
FT TURN 494 495
FT STRAND 498 498
FT STRAND 503 504
FT STRAND 509 512
FT TURN 513 514
FT STRAND 515 518
FT STRAND 1220 1224
FT STRAND 1226 1229
FT TURN 1231 1232
FT STRAND 1234 1242
FT TURN 1246 1247
FT STRAND 1249 1249
FT TURN 1251 1257
FT STRAND 1260 1262
FT STRAND 1263 1270
FT TURN 1272 1273
FT STRAND 1279 1282
FT TURN 1283 1288
FT STRAND 1293 1299
FT TURN 1301 1302
FT STRAND 1306 1307
FT STRAND 1309 1309
FT STRAND 1313 1322
FT TURN 1324 1325
FT STRAND 1329 1342
FT TURN 1344 1345
FT STRAND 1348 1348
FT STRAND 1351 1354
FT STRAND 1356 1360
SQ SEQUENCE 1853 AA, 196831 MW, 3ABDDC03ABFC5372 CRC64;

Query Match 5.3%; Score 91; DB 1; Length 1853;
Best Local Similarity 20.7%; Pred. No. 69;
Matches 62; Conservative 31; Mismatches 119; Indels 88; Gaps 14;

Qy 34 TSNTSSPTQTLSTTKVLRYPDDGPGAPDKDGDGNGPEYIEINLWNILNATGFA 93
Db 353 TINTPTNPANTPVSGN--LKVEP-----YNSNPSETTNSINPKV-----TNTG 395
Qy 94 EMTNLTSGVLHYVQQLDNVLRDRSNVHGYPIFYGNKPWNA-NYATDGPILPKVKS 152
Db 396 SSAIDLKSLTRYTYTVDG--OKDQTEWC-DHAAIIGSNGSYNGITSNVKGTFYKMSST 452
Qy 153 NLTDPLYLTISY--KLBPKNGLPI--NFAIESWLTREAWRTTGINSDEQ----- 196
Db 453 NNAUTYLEISTGTGLBPAGAHVQIQGRFAKDWNQNTQSDNYSFKSASQFVENDQVTAYL 512
Qy 197 -EVMIIWYDGLQAPAGSKV---KEIVVPPIIYNGTFVNATFEVVKANIGWEYVAFRIKTP- 251
Db 513 NGVLAV---GKEFGSVWRPTQVPTTPATTKPPAIT-----KPPA 550
Qy 252 -----IKEGTVPYPAFISVAANISLSP-----NYTELYLEDEVEI 287
Db 551 TTIPSPDDPNAIKIVDTVNAKPGDTVNPVRFSGIPSKGIANCDFVYSYDPNVLEIIEI 610

RESULT 14
APU_THEET
ID APU_THEET STANDARD; PRT; 1481 AA.
AC P38939;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Amylopullulanase precursor (Alpha-amylase/pullulanase) [Includes:
DE Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase);
DE Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase)]

```

```

DE (Alpha-dextrin endo-1,6-alpha-glucosidase)].
GN APU.
OS Thermoanaerobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=1757;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=39E / ATCC 33223;
RX MEDLINE=90147689; PubMed=2302196;
RA Mathupala S.P., Saha B.C., Zeikus J.G.;
RT "Substrate competition and specificity at the active site of
RL amylopullulanase from Clostridium thermohydrosulfuricum.";
CC Biochem. Biophys. Res. Commun. 166:126-132(1990).
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -|- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -|- SUBUNIT: Monomer.
CC -|- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC -|- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; M97665; AAA23201.1; -.
CC PIR; S28669; S28669.
CC RSP; Q08751; 1BVZ.
CC InterPro; IPR006589; Alp_amy1_cat_sub.
CC InterPro; IPR006048; Alpha_amy1_C-
CC InterPro; IPR006047; Alpha_amy1_cat.
CC InterPro; IPR001480; B_lectin.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR004185; Glyco_hydro_13Ig.
CC Pfam; PF02806; alpha-amy1ase_C; 1.
CC Pfam; PF02903; alpha-amy1ase_N; 1.
CC Pfam; PF00041; fn3; 2.
CC SMART; SM00642; Amy; 1.
CC SMART; SM00632; Amy_C; 1.
CC SMART; SM00060; FN3; 2.
CC KW Carbohydrate metabolism; Multifunctional enzyme; Hydrolase;
CC Glycosidase; Calcium-binding; Repeat; Signal.
FT SIGNAL 1 31
FT CHAIN 32 1481
FT DOMAIN 926 1015
FT DOMAIN 1158 1250
FT ACT_SITE 628 628
FT ACT_SITE 657 657
FT ACT_SITE 734 734
FT METAL 398 398
FT METAL 400 400
FT METAL 403 403
FT METAL 404 404
FT METAL 449 449
FT METAL 451 451
FT SEQUENCE 1481 AA; 166362 MW; CB2635960539CAID CRC64;
SQ

Query Match 5.3%; Score 90.5; DB 1; Length 1481;
Best Local Similarity 19.5%; Pred. No. 57;
Matches 52; Conservative 45; Mismatches 91; Indels 79; Gaps 15;

Qy 62 WPGAPDK-----DGDGNPEFYIEINLWNILNATGPAETYNLTSGVLHVQ----- 108

```

Db 95 MEGGVGQGNLSLHSDSVVFFYNNYNTSSVTDSTKYTPPEKUPRIVGTIQSAGA 154  
 QY 109 -----OLDNIVLR-RSNWVHG-----PEIFYGNK-----PMNANYATDGF-----IPL 147  
 Db 155 GDDWKPESTIMRDYKFNWVEYVANYPKRYEFTVLTGSPSWDINYGLNGEQGNIP 214  
 QY 148 -PSKVSNTDFTLTISYKL-----EPKNGL-----P-INFAIESWLTRE- 184  
 Db 215 NVAYDTKTFYVDSYSHNIWTDYNPLPGPNNIYDDLKHTDHPFRFAFGAIGTGT 274  
 QY 185 -AWRTTGINSBOEVMWIIYYDGLQPGSKVKEIIVPIIIVNGTPVNAFVWKANIGWEY 243  
 Db 275 VTLRQARNHDLBSAKI-SYWDIDIKTRTE-----VPMYKIGQSPDGOYEWYVKLSFDY 328  
 QY 244 VAFRIKTP-----IKEGVTTIPYG 262  
 Db 329 -----PTRIYYFILKDGTKTAYYG 348

RESULT 15  
 GUN\_BACSP STANDARD; PRT; 463 AA.  
 AC P29019;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase) (Endo-K).  
 OS Bacillus sp. (strain KSM-330).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1409;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92121380; PubMed=1770347;  
 RA Ozaki K., Sumitomo N., Ito S.;  
 RT "Molecular cloning and nucleotide sequence of the gene encoding an  
 RT endo-1,4-beta-glucanase from Bacillus sp. KSM-330.";  
 RL J. Gen. Microbiol. 137:2299-2305(1991).  
 RN [2]  
 RP SEQUENCE OF 56-75, AND CHARACTERIZATION.  
 RX MEDLINE=91259037; PubMed=2045781;  
 RA Ozaki K., Ito S.;  
 RT "Purification and properties of an acid endo-1,4-beta-glucanase from  
 RT Bacillus sp. KSM-330.";  
 RL J. Gen. Microbiol. 137:41-48(1991).  
 CC -!- FUNCTION: THIS ACID ENDOGLUCANASE IS ACTIVE OVER AN EXTREMELY  
 CC NARROW RANGE OF PH VALUES, BETWEEN 4.5 AND 6.5, WITH AN OPTIMUM  
 CC PH AT 5.2.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- PTM: THE N- AND THE C-TERMINUS MAY BE SUBJECTED TO PROTEOLYSIS.  
 CC -!- MISCELLANEOUS: ONE TRP RESIDUE HAS BEEN PROVED TO BE INVOLVED IN  
 CC THE MECHANISM OF ACTION OF ENDO-K.  
 CC -!- SIMILARITY: Belongs to cellulase family D (family 8 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M68872; AAA22409.1; -;  
 DR PIR; A44808; A44808.  
 DR InterPro; IPR002037; Glyco\_hydro.8.  
 DR InterPro; IPR008928; Glyco\_trans\_6hp.  
 DR Pfam; PF01270; Glyco\_hydro\_8; 1.  
 DR PRINTS; PR00735; GLYDRLASE8.  
 DR PROSITE; PS00812; GLYCOSYL\_HYDROL\_F8; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 27  
 FT SIGNAL 1 27

FT PROPEP 28 55 POTENTIAL.  
 FT CHAIN 56 463 ENDOGLUCANASE.  
 FT ACT\_SITE 130 130 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 191 191 NUCLEOPHILE (POTENTIAL).  
 SQ SEQUENCE 463 AA; 51882 MW; 407FA54F5236C59E CRC64;  
 Query Match 5.3%; Score 90; DB 1; Length 463;  
 Best Local Similarity 19.3%; Pred. No. 14;  
 Matches 46; Conservative 34; Mismatches 58; Indels 100; Gaps 12;  
 QY 91 GFAEMTYNL-----TSGVLHYVQQLDNIVLR-----DR----- 118  
 Db 192 GDLDIAYSLLAHKQWSSGKINYLKEAQNMTTKGKASNVTKNGLNGLDWDGDKSTEDT 251  
 QY 119 ---SNVHGYPEIFY---GNKPNANYATDGPPLPSKVSNTDFTLTISYKLEPKNGL-- 171  
 Db 252 RPSDMMMSHLRAFYEFTGDKTW-----LNVIDNLNTYTNFTNKYSPTKGLIS 299  
 QY 172 -----PINFATESWLTREAWRTTG-----INSDEQEVMIW 201  
 Db 300 DFVVKNPPOPAPKDFLDESKYTDSYYNASRVPLRVMDYAMYGEKRGKVISD--KVATW 357  
 QY 202 I-----YYDGLQPGASKVKE-----IVPIIIVNGTPVNAFVWKANIGWEYV 244  
 Db 358 IKSKTGKNPSKIVDGYKLDGTNIGDYPTAVYVSPFIAAGT-TNSKNQEW-VNSGWDWM 413

Search completed: June 29, 2004, 20:12:05  
 Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 20:06:49 ; Search time 46 Seconds  
(without alignments)  
2188.050 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701

Sequence: 1 MSKKFVIVSILFILLVQAI.....HLEWITNITLPLDRPLIS 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: \*  
1: sp\_archea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriap: \*  
17: sp\_archeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	17 Q9V2T0	Q9V2T0 pyrococcus
2	472	27.7	274	2 P66492	P66492 thermotoga
3	464	27.3	274	16 Q60033	Q60033 thermotoga
4	463.5	27.2	257	2 Q60032	Q60032 thermotoga
5	463.5	27.2	258	16 Q9S5X8	Q9S5X8 thermotoga
6	439.5	25.8	334	17 Q97X08	Q97X08 sulfolobus
7	430	25.3	332	17 Q97VG7	Q97VG7 sulfolobus
8	415.5	24.5	257	2 P96491	P96491 thermotoga
9	415.5	24.4	257	2 Q84828	Q84828 thermotoga
10	221.5	13.0	322	17 Q97V57	Q97V57 sulfolobus
11	143.5	8.4	261	2 Q7X494	Q7X494 bacillus 11
12	128	7.5	260	2 Q33897	Q33897 rhodothermu
13	114.5	6.7	381	2 Q54331	Q54331 streptomyce
14	111.5	6.6	269	16 Q82K36	Q82K36 streptomyce
15	111	6.5	239	3 Q74705	Q74705 aspergillus
16	111	6.5	1289	17 Q96XM6	Q96XM6 sulfolobus

17	110.5	6.5	264	2	Q31030	Q31030 pectobacter
18	109	6.4	304	16	Q9EP57	Q9EP57 oceanobacil
19	108	6.3	371	2	Q9KIH1	Q9KIH1 streptomyce
20	107.5	6.3	381	16	Q9RJY3	Q9RJY3 streptomyce
21	107	6.3	1612	11	Q9WUT8	Q9WUT8 mus musculu
22	104.5	6.1	1062	17	Q96X97	Q96X97 sulfolobus
23	103.5	6.1	382	2	Q59963	Q59963 streptomyce
24	103	6.1	434	16	Q7WLF6	Q7WLF6 bordetella
25	103	6.1	859	5	Q21401	Q21401 caenorhabdi
26	102.5	6.0	533	5	Q18684	Q18684 caenorhabdi
27	101	5.9	713	5	Q19687	Q19687 caenorhabdi
28	100.5	5.9	534	12	Q67668	Q67668 goose parvo
29	100.5	5.9	587	12	Q67667	Q67667 goose parvo
30	100.5	5.9	732	12	Q67666	Q67666 goose parvo
31	100	5.9	263	16	Q9AN41	Q9AN41 bradyrhizob
32	100	5.9	597	17	Q97I01	Q97I01 sulfolobus
33	99.5	5.8	1041	16	Q8EU25	Q8EU25 mycoplasma
34	99.5	5.8	1043	16	Q8EV01	Q8EV01 mycoplasma
35	99	5.8	320	10	Q9ATK2	Q9ATK2 arabidopsis
36	99	5.8	384	2	Q8X602	Q8X602 streptomyce
37	99	5.8	620	17	Q97UF5	Q97UF5 sulfolobus
38	99	5.8	660	5	Q86K14	Q86K14 dictyosteli
39	99	5.8	742	11	Q8BNV3	Q8BNV3 mus musculu
40	99	5.8	2029	17	Q8TI44	Q8TI44 methanosarc
41	98.5	5.8	390	5	Q86JF6	Q86JF6 dictyosteli
42	98.5	5.8	561	16	Q8NY14	Q8NY14 staphylococ
43	98.5	5.8	572	16	Q9X20	Q9X20 staphylococ
44	97.5	5.7	279	10	Q49899	Q49899 medicago sa
45	97.5	5.7	377	2	O08468	O08468 streptomyce

#### ALIGNMENTS

#### RESULT 1

Q9V2T0 PRELIMINARY; PRT; 319 AA.  
ID Q9V2T0;  
AC Q9V2T0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Endoglucanase A precursor (Endo-1,4-beta-glucanase b).  
GN EGLA OR PF0854.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99084968; PubMed=9864341;  
RA Bauer M.W., Driskill L.E., Callen W., Snead M.A., Mathur B.J.,  
RA Kelly R.M.;  
RT "An endoglucanase, EglA, from the hyperthermophilic archaeon  
RT Pyrococcus furiosus hydrolyzes beta-1,4 bonds in mixed-linkage  
RT (1--3), (1-->4)-beta-D-glucans and cellulose.";  
RL J. Bacteriol. 181:284-290(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb P.T., Brown J.R.;  
RA "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR ENBL; AF181032; AAD54602.1; -  
DR ENBL; AE010200; AAL80978.1; -  
DR GO; GO:000810; F:cellulase activity; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR008985; ConA like leg gl.  
DR InterPro; IPR002594; Glyco\_hydro\_12.  
DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
KW Signal; Complete proteome.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 319 ENDOGLUCANASE A.

SQ SEQUENCE 319 AA; 35988 MW; 51686BCA569F7A43 CRC64;  
Query Match 100.0%; Score 1701; DB 17; Length 319;  
Best Local Similarity 100.0%; Pred. No. 3.3e-128;  
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSKKFVIVSILTIILLVQAIYVEKYHTSEDKSTNTSTPTTLSTTKVLYKIRYDDG 60  
DB 1 MSKKFVIVSILTIILLVQAIYVEKYHTSEDKSTNTSTPTTLSTTKVLYKIRYDDG 60  
QY 61 EWPAGPIDKGDGNPEFIEINLNLNATGFAEMTYNLTSGVLHYVQOLNIVLRDSN 120  
DB 61 EWPAGPIDKGDGNPEFIEINLNLNATGFAEMTYNLTSGVLHYVQOLNIVLRDSN 120  
QY 121 VWHGYPEIFYGNKPNWANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
DB 121 VWHGYPEIFYGNKPNWANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180  
QY 181 LTREAWRTGINSDEQVMIWYDGLQPGAGSKYKEIIVPIIVNGTTPVNAATFEVWKANIG 240  
DB 181 LTREAWRTGINSDEQVMIWYDGLQPGAGSKYKEIIVPIIVNGTTPVNAATFEVWKANIG 240  
QY 241 WEYVAFRIKTPKEGTVPYPGAFISVAANISSLPNTYLYLEDVEIGTEFGTPTTSAH 300  
DB 241 WEYVAFRIKTPKEGTVPYPGAFISVAANISSLPNTYLYLEDVEIGTEFGTPTTSAH 300  
QY 301 LEWITNLTPLDRPLIS 319  
DB 301 LEWITNLTPLDRPLIS 319  
RESULT 2  
P36492  
ID P36492 PRELIMINARY; PRT; 274 AA.  
AC P36492;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Endo-1,4-beta-glucanase B (EC 3.2.1.4) (Cellulase)  
DE (Carboxymethyl cellulase).  
GN CELB.  
OS Thermotoga neapolitana.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2337;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bok J.D., Eveleigh D.E.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-137 FROM N.A.  
RC STRAIN=Z2706-MC24;  
RA Zverlov V.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-137 FROM N.A.  
RC STRAIN=Z2706-MC24;  
RX MEDLINE=93356813; PubMed=8352795;  
RA Dakhova O., Kurepina N., Zverlov V., Svetlichnyi V.,  
RA Velikodvorskaya G.;  
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana  
genes coding for enzymes of carbohydrate substrate degradation";  
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).  
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
LINKAGES IN CELLULOSE.  
DR EMBL; U93354; AAC95060.1; -;  
DR EMBL; Z86103; CAB06782.1; -;  
DR GO; GO:0008810; P:cellulase activity; IEA.  
DR GO; GO:0016798; P:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR000985; ConA like lec.gl.  
DR InterPro; IPR002594; Glyco\_hydro\_12.  
DR Pfam; PF01670; Glyco\_hydro\_12; 1.

DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 274 AA; 31752 MW; 6B21C4P8CA7C9AA2 CRC64;  
Query Match 27.7%; Score 472; DB 2; Length 274;  
Best Local Similarity 40.4%; Pred. No. 8.4e-30;  
Matches 93; Conservative 47; Mismatches 80; Indels 10; Gaps 4;  
QY 79 IEINLNLNATGFAEMTYNLTSGVLHYVQOLNIVLRDSNHWGYPYFIEFYGNKPNW 138  
DB 39 MELFNWVKSIEG-ETWLKFGKQVYADIVNIVLQNPDSWVHGYPYFIEFYGNKPNW 96  
QY 139 YATDGPIDPSPKVSNTLDFYLTISYKLEPKNGLPINFAIESWLTREAWRTTGINSDQEV 198  
DB 97 --NSGTEILFVKVQKLPDFYVTLDSIYVNDLPINLAMETWITRKPQTS-VSSGDVEI 153  
QY 199 MIWIVYDGLQPGAGSKYKEIIVPIIVNGTTPVNAATFEVWKANIGWEYVAFRIKTPKEGT 258  
DB 154 MWYFYNILFPGQKVDFTTIEINGSPVETKMDVYPAPWGDVLAFLTTPMDGRVK 213  
QY 259 IPYGAFTSVAANI-----SSLPNTYLYLEDVEIGTEFGTPTTSAHLEW 303  
DB 214 FNVKDFEKAEEVIKGSRTVENFDEMPCVWEIGTEFGDPNTTAAKFGW 263  
RESULT 3  
Q60033 PRELIMINARY; PRT; 274 AA.  
AC Q60033;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Endo-1,4-beta-glucanase precursor (EC 3.2.1.4) (Cellulase)  
DE (Endoglucanase) [Carboxymethyl cellulase].  
GN CELB OR TMI525.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8;  
RX MEDLINE=96425879; PubMed=8828221;  
RA Liebl W., Ruile P., Bronnenmeier K., Riedel K., Lottspeich F.,  
RA Greif I.;  
RT "Analysis of a Thermotoga maritima DNA fragment encoding two similar  
thermostable cellulases, Cella and CelB, and characterization of the  
recombinant enzymes";  
RL Microbiology 142:2532-2542(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima";  
RL Nature 399:323-329(1999).  
DR EMBL; Z69341; CAA93274.1; -;  
DR EMBL; AEC01800; AAD36592.1; -;  
DR FIR; A72241; A72241.  
DR TIGR; TMI525; -;  
DR GO; GO:0008810; P:cellulase activity; IEA.  
DR GO; GO:0016798; P:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR008985; ConA like lec.gl.  
DR InterPro; IPR002594; Glyco\_hydro\_12.  
DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
DR ProDom; PD004316; Glyco\_hydro\_12; 1.



KW Signal; Hydrolase; Glycosidase; Complete proteome.  
 FT SIGNAL 17 POTENTIAL.  
 FT CHAIN 18 274 ENDO-1,4-BETA-GLUCANASE.  
 SQ SEQUENCE 274 AA; 31734 MW; 720A5ACAD7BA358 CRC64;  
 Query Match 27.3%; Score 464; DB 16; Length 274;  
 Best Local Similarity 40.4%; Pred. No. 3.7e-29;  
 Matches 93; Conservative 43; Mismatches 84; Indels 10; Gaps 4;  
 QY 79 IEINLNLNATGFAEMTYNLTSGVLHYVQOQDNVILDRSNWVHGYPFIYGNKEPNAN 138  
 DB 39 MELFNWVKVYEG--ETWLKDEGEVFFADLYNLVLPNDVSWVHGYPFIYGYKQWAGH 96  
 QY 139 YATDGPILPSKVSNTDFLTTSYKLEPKNGLPINPAIESWLTREAWRTTGINSDEQEV 198  
 DB 97 --NSGVFEFLPVKVDLPDFVTLDSIYENNLPLINAMETWITRSPDQTS--VSSGDAHI 153  
 QY 199 MIWYDGLQAGSKVKEIVVPIVNGTVPNATFEVWKNIGVEYAFRIKTIPEKGTVT 258  
 DB 154 MVWFYNNVLPFGQKVDEFTTVEINGVKQETKWDVYFAPWGDYLAFLRTTPMKEGVKV 213  
 QY 259 IPYGAFISVAANI-----SSLPNYTELYLEDVIGTEFGPPTSAHLEW 303  
 DB 214 INVKDFYKAAEVVKKHSTRIDNPEELYFCVMBIGTEFGDPNTAAKFGW 263  
 RESULT 4  
 ID Q60032 PRELIMINARY; PRT; 257 AA.  
 AC Q60032;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Endo-1,4-beta-glucanase precursor (EC 3.2.1.4).  
 GN CELA.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.;  
 RL Nature 399:323-329 (1999).  
 DR EMBL; AE001800; AAC36591.1; -.  
 DR PIR; H72240; H72240.  
 DR TIGR; TM1524; -.  
 DR GO; GO:0008810; F:cellulase activity; IEA.  
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR008985; ConA-like\_lect\_g1.  
 DR InterPro; IPR002594; Glyco\_hydro\_12.  
 DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 257 AA; 29733 MW; D05CB07EBB9FB36B CRC64;  
 Query Match 27.2%; Score 463.5; DB 2; Length 257;  
 Best Local Similarity 34.8%; Pred. No. 3.7e-29;  
 Matches 89; Conservative 53; Mismatches 95; Indels 19; Gaps 5;  
 QY 59 DGEWPGAPIDKGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOQDNVILDR 118  
 DB 11 DFWNGIPLS-----MELNLWNKEYSGSVAMKFD--GEKIFDADIQNLSPKEP 58  
 QY 119 SNWVHGYPEIFYGNKPNWANYATDGPILPSKVSNTDFLTISYKLEPKNGLPINFAIE 178  
 DB 59 ERYVLGYPEFYGYKPEWNTAEGSKLPV--VSSMKSFSEVSVFDIHHEPSLPLNFAME 116  
 FT SIGNAL 257 POTENTIAL.  
 SQ SEQUENCE 257 AA; 29733 MW; D05CB07EBB9FB36B CRC64;  
 Query Match 27.2%; Score 463.5; DB 2; Length 257;  
 Best Local Similarity 34.8%; Pred. No. 3.7e-29;  
 Matches 89; Conservative 53; Mismatches 95; Indels 19; Gaps 5;  
 QY 59 DGEWPGAPIDKGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOQDNVILDR 118  
 DB 11 DFWNGIPLS-----MELNLWNKEYSGSVAMKFD--GEKIFDADIQNLSPKEP 58  
 QY 119 SNWVHGYPEIFYGNKPNWANYATDGPILPSKVSNTDFLTISYKLEPKNGLPINFAIE 178  
 DB 59 ERYVLGYPEFYGYKPEWNTAEGSKLPV--VSSMKSFSEVSVFDIHHEPSLPLNFAME 116  
 QY 179 SWLTREAWRTTGINSDEQEVMIWYDGLQAGSKVKEIVVPIVNGTVPNATFEVWKN 238

DB 117 TWLTREKYQTEASIGD-VEIMVWFYFNLTGGKIEFTIPFVLNGESVEGTWELWLA 175  
 QY 239 IGWEYVAFRIKTIPEKGTVTIPYGAFFISVAANIS----LPNYTELYLEDVIGTEFGTP 294  
 DB 176 WGDYLAFLKDPVKVGRVKFDVRHFLDAAGKALSSARVKDFEDLYFTVWVEIGTEFGSP 235  
 QY 295 STTSAHLEWITNITL 310  
 DB 236 ETKSAQFGWKFNFSI 251  
 RESULT 5  
 ID Q9S5X8 PRELIMINARY; PRT; 258 AA.  
 AC Q9S5X8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Endoglucanase.  
 GN TM1524.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.;  
 RL Nature 399:323-329 (1999).  
 DR EMBL; AE001800; AAC36591.1; -.  
 DR PIR; H72240; H72240.  
 DR TIGR; TM1524; -.  
 DR GO; GO:0008810; F:cellulase activity; IEA.  
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR008985; ConA-like\_lect\_g1.  
 DR InterPro; IPR002594; Glyco\_hydro\_12.  
 DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 258 AA; 29833 MW; DS15E14EE2417A6B CRC64;  
 Query Match 27.2%; Score 463.5; DB 16; Length 258;  
 Best Local Similarity 34.8%; Pred. No. 3.7e-29;  
 Matches 89; Conservative 53; Mismatches 95; Indels 19; Gaps 5;  
 QY 59 DGEWPGAPIDKGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOQDNVILDR 118  
 DB 12 DFWNGIPLS-----MELNLWNKEYSGSVAMKFD--GEKIFDADIQNLSPKEP 59  
 QY 119 SNWVHGYPEIFYGNKPNWANYATDGPILPSKVSNTDFLTISYKLEPKNGLPINFAIE 178  
 DB 60 ERYVLGYPEFYGYKPEWNTAEGSKLPV--VSSMKSFSEVSVFDIHHEPSLPLNFAME 117  
 QY 179 SWLTREAWRTTGINSDEQEVMIWYDGLQAGSKVKEIVVPIVNGTVPNATFEVWKN 238  
 DB 118 TWLTREKYQTEASIGD-VEIMVWFYFNLTGGKIEFTIPFVLNGESVEGTWELWLA 176  
 QY 239 IGWEYVAFRIKTIPEKGTVTIPYGAFFISVAANIS----LPNYTELYLEDVIGTEFGTP 294  
 DB 177 WGDYLAFLKDPVKVGRVKFDVRHFLDAAGKALSSARVKDFEDLYFTVWVEIGTEFGSP 236  
 QY 295 STTSAHLEWITNITL 310  
 DB 237 ETKSAQFGWKFNFSI 252

```

RESULT 6
Q97X08 PRELIMINARY; PRT; 334 AA.
AC Q97X08
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4).
GN SSO1949.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awey M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Etrauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AB006749; AAK41590.1; -
DR PIR; G90291; G50291.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD04316; Glyco_hydro_12; 1.
DR KW Hydrolase; Glycosidase; Complete proteome.
SQ SEQUENCE 334 AA; 37619 MW; CB3F6139363934A8 CRC64;

Query Match 25.8%; Score 439.5; DB 17; Length 334;
Best Local Similarity 33.4%; Pred. No. 4.4e-27;
Matches 117; Conservative 57; Mismatches 121; Indels 55; Gaps 17;

QY 1 MSKKKFFVIVSLTLLV-----QAIYFVKYHTSBDKSTN-----TSSTP 41
DB 3 MKKLYIIVPIVIVGVIGALYL---HQSPVKTSITVTNTTILMSITNTVP 59
QY 42 PQTLLSTTKVLRYPDDGWPAGAPDKDGNPEFYGNKPNANYATDGPILPSKVSNLT 101
DB 60 TTVTPTTSSIPQLIYVTSASSPTPVYLNSTVSPFYLEVMNNAKTMNGNYTMVFNPLT 119
QY 102 GVLHVVOOLDNIVLRDRSNVHGYPEIFYGNKPNANYATDGPILPSKVSNLT---DF 157
DB 120 RTLVSFNLTVQV---NPLQWTNGYPEIYVGRKPDWTSYA--GNI-FPMRIGNMTPPFWVSF 173
QY 158 YLTISYKLEPKNGLPINF--AIESWLTREAWR---TTGINSDEQVMIWYDGLQAPAGS 212
DB 174 YINLT-KLDPS---INFDIASDAWIVRPOIAFSPGAPGNGDIEIMWLFQNLQAPAG 228
QY 213 KVKIIVPIVIVNGTVPVNATEVWK-ANI---GWEYVAFRIKTP-----IKEGTVTIPYGA 264
DB 229 QVGEVPIPIYINHTLVNATFQVWKKNVPMGWEYIAFR---PDGKVTNGYVAYEPNLF 285
QY 265 ISVAANISSLPNYTELYLDEVEIGTEFGT-PTSTSALHLEWITNITLTP 313
DB 286 IXLNNPASY-NITNYLTDWERTGTWTSNGTAYFSWTSINFTYELL 334

RESULT 7
Q97YG7 PRELIMINARY; PRT; 332 AA.
AC Q97YG7
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase (EC 3.2.1.4).
GN CELA.

```

```

DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4).
GN SSO1354.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awey M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Etrauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AB006749; AAK41590.1; -
DR PIR; G90291; G50291.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD04316; Glyco_hydro_12; 1.
DR KW Hydrolase; Glycosidase; Complete proteome.
SQ SEQUENCE 332 AA; 37295 MW; F896165AE426B9E CRC64;

Query Match 25.3%; Score 430; DB 17; Length 332;
Best Local Similarity 32.2%; Pred. No. 2.5e-26;
Matches 111; Conservative 60; Mismatches 124; Indels 50; Gaps 15;

QY 4 KKFVIVSLTLLV-----QAIYFVKYHTSBDKSTN-----PQTLL 46
DB 3 KLYIVLPVIVIAIGVMGGIYHQQSLSVKRPVTFSTTSTTTTNAITTTVTQVT 62
QY 47 STTKVLKIRYPDDGWPAGAPDKDGNPEFYGNKPNANYATDGPILPSKVSNLT 106
DB 63 STTSVQLIYVTSASSPTPVYLNSTVSPFYLEVMNNAKTMNGNYTMVFNPLTSLV 122
QY 107 VOOLDNIVLRDRSNVHGYPEIFYGNKPNANYATDGPILPSKVSNLT---DFYLTIS 162
DB 123 SFNLQV---KPLEWTNGYPEIYVGRKPDWTSYA--GNI-FPMRIGNMTPPWVSFYINLT 176
QY 163 YKLEPKNGLPINF--AIESWLTREAWR---TTGINSDEQVMIWYDGLQAPAGSKYKEI 217
DB 177 KLDPS---INFDIASDAWIVRPOIAFSPGAPGNGDIEIMWLFQNLQAPAGQVGV 231
QY 218 VVPIVINGTVPVNATEVWKANI---GWEYVAFRIKTP-----IKEGTVTIPYGA 269
DB 232 VVPIVINGTVPVNATEVWKANI---GWEYVAFRIKTP-----IKEGTVTIPYGA 288
QY 270 NISSLPNTELYLDEVEIGTEFGT-PTSTSALHLEWITNITLTP 313
DB 289 NFTSY-NITNYLTDWERTGTWTSNGTAYFSWTSINFTSEILL 332

RESULT 8
P96491 PRELIMINARY; PRT; 257 AA.
AC P96491
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase (EC 3.2.1.4).
GN CELA.

```

```

OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
CX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2706-MC24;
RA Zverlov V.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2706-MC24;
RX MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kurepina N., Zverlov V., Svetlichnyi V.,
RA Velikodvorskaya G.;
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana
RT genes coding for enzymes of carbohydrate substrate degradation.";
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
DR EMBL; Z86103; CAB06783.1; -.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 257 AA; 30192 MW; D6E148CABAC864F7 CRC64;

Query Match 24.5%; Score 416.5; DB 2; Length 257;
Best Local Similarity 33.6%; Pred. No. 2.2e-25;
Matches 86; Conservative 51; Mismatches 100; Indels 19; Gaps 6;

QY 59 DGEWPGAIKDKGNGNPEFYIENLNLNATGPAETYNLTSGVLHYVQQLDNIVLRDR 118
DB 11 DFRNDMPLS-----MELNLNMYRYTGTVMRFD--GERLTFNGDVEDLSAREP 58

QY 119 SNWVHGYPEIFYGNKPNMANYATDGPILPSPKVSNTLDFYLTISYKLEPKNGLPINFAIE 178
DB 59 ERYILGYPEFYGYKPEW-RHAAEG-TKLPLVSSVESFTVELSFEIDHMPSLPLNFAME 116

QY 179 SWLTREAWRTTGINSDEQVMIWYDGLQAGSKVEIIVPIVNGTPVYNATFEVWKAN 238
DB 117 TWLTREKYQVEASIGD-VEIMVWFYFNELTPGGKKVGEYTVSFELNGEHKGIWELWHA 175

QY 239 IGWEYVAFRIKTIKEGTVTIPYGAFISVAANISS-----LPNYTELYLEDVEIGTEFGTP 294
DB 176 WNWYDLAFRLKNPKVRGRVRFNVKDFLDVAGEYLSRSTRVKDFDLYFTVWEIGTEFGSP 235

QY 295 STTSAHLEWITNITL 310
DB 236 ETKSARFGWTNNFSI 251

RESULT 9
ID O08428 PRELIMINARY; PRT; 257 AA.
AC O08428;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase A (EC 3.2.1.4).
GN CELA.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
CX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054912; PubMed=9835561;
RA Bok J.D., Vernool D.A., Eveleigh D.E.;
RT "Purification, characterization, and molecular analysis of
RT thermostable cellulases CELA and CelB from thermotoga neapolitana.";
RL Appl. Environ. Microbiol. 64:4774-4781(1998).

```

```

DR EMBL; U93354; AAC95059.1; -.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR InterPro; IPR002594; Glyco_hydro_12; 1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 257 AA; 30201 MW; 14856B9E8AD0EDEF CRC64;

Query Match 24.4%; Score 415.5; DB 2; Length 257;
Best Local Similarity 33.6%; Pred. No. 2.6e-25;
Matches 86; Conservative 51; Mismatches 100; Indels 19; Gaps 6;

QY 59 DGEWPGAIKDKGNGNPEFYIENLNLNATGPAETYNLTSGVLHYVQQLDNIVLRDR 118
DB 11 DFRNDMPLS-----MELNLNMYRYTGTVMRFD--GERLTFNGDVEDLSAREP 58

QY 119 SNWVHGYPEIFYGNKPNMANYATDGPILPSPKVSNTLDFYLTISYKLEPKNGLPINFAIE 178
DB 59 ERYILGYPEFYGYKPEW-RHAAEG-TKLPLVSSVESFTVELSFEIDHMPSLPLNFAME 116

QY 179 SWLTREAWRTTGINSDEQVMIWYDGLQAGSKVEIIVPIVNGTPVYNATFEVWKAN 238
DB 117 TWLTREKYQVEASIGD-VEIMVWFYFNELTPGGKKVGEYTVSFELNGEHKGIWELWHA 175

QY 239 IGWEYVAFRIKTIKEGTVTIPYGAFISVAANISS-----LPNYTELYLEDVEIGTEFGTP 294
DB 176 WNWYDLAFRLKNPKVRGRVRFNVKDFLDVAGEYLSRSTRVKDFDLYFTVWEIGTEFGSP 235

QY 295 STTSAHLEWITNITL 310
DB 236 ETKSARFGWTNNFSI 251

RESULT 10
ID Q97VS7 PRELIMINARY; PRT; 322 AA.
AC Q97VS7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase (Endo 1,4 beta glucanase), putative (celB
DE (EC 3.2.1.4).
GN CELB OR CELS OR SSO2534.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
CC Sulfolobus.
CX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic V., Allard G.,
RA Awaysz M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozeva C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49255 / DSM 5833 / MT-4;
RX Limauro D., Fiorentino G., Cannio R., Rossi M., Bartolucci S.;
RT "Identification and molecular characterization of thermophilic
RT endoglucanase gene, celS, from Sulfolobus solfataricus.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS006850; AAK42663.1; -.
DR EMBL; AJ296029; CAC39483.1; -.

```

```

DR PIR: H90425; H90425.
DR GO: GO:0008810; F:cellulase activity; IEA.
DR GO: GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro: IPR008985; ConA like lec g1.
DR Pfam: PF01670; Glyco_hydro_12.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR HydroLase: Glycosidase; Complete proteome.
SQ SEQUENCE 322 AA; 36710 MW; 6C32D7765173A768 CRC64;

Query Match
Best Local Similarity 27.3%; Pred. No. 1.2e-09; Length 322;
Matches 89; Conservative 52; Mismatches 132; Indels 53; Gaps 17;

QY 5 KEVIVSIITILLVQAIYVE--KYH--TSEDKSTNTSSTP-----PQTLSTTKVLKIRY 56
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KLIPFIVVVIIVIGIIVSIEFGKHFQNASLRTSTERFLFAHNRPFSL----- 52

QY 57 PDGSGWPCAPDK-----DGDGNPEFYIEINLWNLINATGFAEMTYNLTSGVLHVQQLDN 112
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 ---GNYSNSADALAINSSNATLWSPFLNIGYALGNVMTINI--NYLHVAINLSQ 107

QY 113 IVLDRSNWVGYPFIYGNXP--NNAYATDGP--IPLPSKVSNLTPYLTISYKLEPKN 169
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 I-SKISNVVDGYPGLMYGQELWPFMYRTTQLQFLSLPMIVLRLPNFYSLNYSVYLIN 166

QY 170 GLPINFATESMLTREAMTTGINSDEQEWMIYYD---GLQAGSKVKEIWPVLIYNGT 226
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 GSIDFSYDILWSONP-NITLQYGDFFIMIMWYNENLSHTPYFIYGNMSIFTLNGK 225

QY 227 PVNATFEVW-----KANIGWEYVAF--RIKTIKEGTVTIPY-----GAFISVA-ANIS 272
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 IENLSWEVVLPRGTSAN-GWTGVVFLSLPEKPAFGVPIGYILKMGSVIERAGVNIY 284

QY 273 SLPNTELYLEDVEIGTEFGTPSTIS 298
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 NVNTY---YLDAIQGMFEFSDNQGTA 307

RESULT 11
QYX4S4 PRELIMINARY; PRT; 261 AA.
ID Q7X4S4
AC Q7X4S4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-beta-1,4-glucanase (EC 3.2.1.4).
GN CEL12A.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GXN151.
RA Liu Y., Feng J., Duan C., Zhao G., Bai X., Zhang C., Tang J., Ma Q.;
RT "Sequencing of cellulase gene cel12A of Bacillus licheniformis strain
GXN151 and its over-expression in Escherichia coli."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY291066; AAP44491.1; -.
DR EMBL; AY291066; AAP44491.1; -.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 261 AA; 29068 MW; 444CA54F697D95C3 CRC64;

Query Match
Best Local Similarity 8.4%; Score 143.5; DB 2; Length 261;
Matches 63; Conservative 37; Mismatches 97; Indels 67; Gaps 15;

QY 66 PIDGDGNGPEFYIEINLWNLINATGFAEMTYNLTSGVLHVQQLDNVLDRSNW---- 121
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35 PSDKLYFKKKYKYLIFNNVWGDQVSGWQTIYHNSDSMGVW-----NWPST 83

QY 122 --VHGYPEIFYGNKPNAYATDGPILPSKVSNLTPYLTISYKLEPKNGLPINFAIES 179

```

```

Db 84 STVKAYPSIVSG-WHTEGYTAGS--GPFTRLSDDQKNTKYSYIS-ANG-TYNAAYDI 138
QY 180 WL---TREAWRTTGINSDEQEWMIYYDGLQAPAGSKVKEIWPVLIYNGTVPNATFEVWK 236
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 WLHNTNKASWD---SAPTDEIMWLNTNAGPAGSYVETVSGI-----GHSWKVYK 186

QY 237 ANI-----GWEVVAPRIKTIKEGTVTIPYCAFISVAANISSLPNY-----TELY 281
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 GYIDAGGGKGNVFSF-IRANTQ-----SANLINRDTNYLADSKQWLKSKTY 234

QY 282 LEDVEIGTE-FGTPSTTSAHLEW 304
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 VSSVEFGTEVFG-GTQINISNW 256

RESULT 12
QYX3897 PRELIMINARY; PRT; 260 AA.
ID Q33897
AC Q33897;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase (EC 3.2.1.4).
GN CELA.
OS Rhodothermus marinus (Rhodothermus obamensis).
OC Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC Crenotrichaceae; Rhodothermus.
OX NCBI_TaxID=29549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ITI378;
RX MEDLINE=98242392; PubMed=9581231;
RA Halldorsdottir S., Thorolfssdottir E.T., Spilliaert R., Johansson M.,
RA Thorbjarnardottir S.H., Paladottir A., Hreggvidsson G.O.,
RA Kristjansson J.K., Holst O., Eggertsson G.,
RT "Cloning, sequencing and overexpression of a Rhodothermus marinus gene
encoding a thermostable cellulase of glycosyl hydrolase family 12."
RL Appl. Microbiol. Biotechnol. 49:277-284 (1998).
DR EMBL; U72637; AAB65594.1; -.
DR GO: GO:0008810; F:cellulase activity; IEA.
DR GO: GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro: IPR008985; ConA-like lec g1.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR PROSITE; PS00013; PROKAE_LIPOPROTEIN; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 260 AA; 28770 MW; 94F197DB7D0D247A CRC64;

Query Match
Best Local Similarity 7.5%; Score 128; DB 2; Length 260;
Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;

QY 59 DGEWGPATDKDGDNGNPEFYIEI-----NLWNLINATGFAEMTYNLTSGVLHV 107
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 19 DWLFPDGDNGKEPEPEPEPTVELCGRWDRDVGGRYRVINNVWGAETAQCIIEVGL---- 74

QY 108 QOLDNLVL-----RDRSNWVGYPEIFYGNKPNAYATDGPILPSKVSNLTPYLTISY 163
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 -ETGNFTITRADHDNGNVAAVPAIYFGCH-WAPARATRDCAARAGAVRAHELDVT--- 129

QY 164 KLEPKNGLPI-----NFAIESMLTREAMTTGINSDEQEWMIYYD-GLQAGSKVKEI 217
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 -----PITGRWNAAYDWFSPVTSNGY-SGGAELMWLNWGNVMPGGSRVATV 180

QY 218 VVPIIYNGTVPNATFEVWKANIGWEYVAFRIKTIKEGTVTIPYCAFISVAANISSLPNY 277
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 EL-----ACATVEVWVADWNNYIAYRRTTP-TTTSVELDLKAFITDVAAGYI--R 229

```

QY 278 TELYLEDEVEIGTE 290  
Db 230 PEWYLHAVETGFE 242

RESULT 13

Q54331 PRELIMINARY; PRT; 381 AA.

AC Q54331, 1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Cellulase B precursor.  
GN CELB

OS Streptomyces lividans.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1916;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=66;  
RX MEDLINE=94288649; PubMed=9017952;  
RA Wittmann S., Sharek F., Kluepfel D., Morosoli R.;  
RT "Purification and characterization of the CelB endoglucanase from  
Streptomyces lividans 66 and DNA sequence of the encoding gene.";  
RL Appl. Environ. Microbiol. 60:1701-1703(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=66;  
RA Sharek F.;  
RL EMBL; U04629; AAB71950.1; -;  
DR PDB; 1NLR; 25-NOV-98.  
DR PDB; 2NLR; 07-FEB-00.  
DR GO; GO:0008810; F:cellulase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR001919; Bac\_celose-bind.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR008965; ConA like lec gl.  
DR InterPro; IPR002594; Glyco\_hydro\_12.  
DR Pfam; PF00553; CEM\_2; 1.  
DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
DR SMART; SM00637; CBD II; 1.  
DR PROSITE; PS00561; CBD\_BACTERIAL; 1.  
KW Signal.  
FT SIGNAL.  
FT CHAIN.  
SQ SEQUENCE 381 AA; 39239 MW; AYE99BF590FA24EC CRC64;

Query Match 6.7%; Score 114.5; DB 2; Length 381;  
Best Local Similarity 27.1%; Pred. No. 0.58;  
Matches 52; Conservative 21; Mismatches 80; Indels 39; Gaps 10;

QY 125 YPEIPGNKPNVATDGPPLPSKVSNTDFLTISYKLEPKNGLPINFAESWLTRE 184  
Db 97 YPSVFNGCHYTNCSFGTD----LPVRLDTVAAPSSISYGF--VDGAVYNASYDIWLDPT 150  
QY 185 AWRTTGINSDEQVNMWITYDQ--LQPSGSKYKEIIVPIIVNGTPV--NATPEVW----- 235  
Db 151 A-RTDGVN--QTEINIMFNVRGPIQIGSPV-----GTASVCGRTWEVWSGGNS 197  
QY 236 -----KANIGHEY-VAPRIKTIPEKGVTTIPYGAFISVAANISLNPNTELYLEDV 285  
Db 198 NDVLSFVAFSAISGWSFDVMDFVATVARGLAENDW-YLTSVQAGFPWQNGAGLVNSF 256  
QY 286 EIGTEFGTPTST 297  
Db 257 SSTVETGPGGT 268

RESULT 14

Q82K36 PRELIMINARY; PRT; 269 AA.

AC Q82K36;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative endo-1,4-beta-glucanase.  
GN CELA4 OR SAV2568.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing Streptomyces  
RT metabolites.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis";  
RL Nat. Biotechnol. 21:526-531(2003).  
DR EMBL; AP005031; BAC70279.1; -;  
DR GO; GO:0008810; F:cellulase activity; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR008965; ConA like lec gl.  
DR InterPro; IPR002594; Glyco\_hydro\_12.  
DR Pfam; PF01670; Glyco\_hydro\_12; 1.  
DR ProDom; PD004316; Glyco\_hydro\_12; 1.  
KW Complete proteome.  
SQ SEQUENCE 269 AA; 29362 MW; 07B90B00F577BC5F CRC64;

Query Match 6.6%; Score 111.5; DB 16; Length 269;  
Best Local Similarity 22.0%; Pred. No. 0.64;  
Matches 51; Conservative 38; Mismatches 96; Indels 47; Gaps 10;

QY 76 EPIEINLNLNATGFAETMYNTLTSGVLHYVQQLDNIVLRDRSNWV-----HGYEIPY 130  
Db 50 KYVNNLWQDKATGTCTQCVWDDSRSG-----STISWGTDSWANNAGHDYDVKSY 101  
QY 131 GNK-----PWANYATDGPILPSKVSNTDFLTISYKLEPKNGLPINFAESWLTRE 184  
Db 102 ASSVLGHWGKADKAATG---LPVVGDRKSVRTTWDFSVGAGPG--TNVAYDLWLHAK 157  
QY 185 AWRTTGINSDEQ-----EVMWMT-VYDGLQPSGSKYKEIIVPIIVNGTPVNTATFEVWKN 238  
Db 158 N-----NADWQDPDEIMWLNROGGAGPLGTYGVSLSG-----GAMWDIYQGD 203  
QY 239 IGHEYVAFRIKTIPEKGVTTIPYGAFISVAANISLNPNTELYLEDVETGTE 290  
Db 204 IGKVIYSFVRTTNTTKATLDLDDFTQALVRRKLLS---DDKYVSGIEAGTE 251

RESULT 15

C74705 PRELIMINARY; PRT; 239 AA.

AC C74705;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Endoglucanase A precursor (EC 3.2.1.4) (CELLULOSE)  
DE (Endo-1,4-beta-glucanase A) (CARBOXYMETHYLCELLULOSE).  
GN EGLA OR CEKA.

